Universität Basel
Wirtschaftswissenschaftliche Fakultät

May 2018

Estimating Survival Times Using Swiss Hospital Data

WWZ Working Paper 2018/14

Florian Kuhlmey, Matthias Minke

A publication of the Center of Business and Economics (WWZ), University of Basel.

Universität Basel
Peter Merian-Weg 6
4052 Basel, Switzerland
wwz.unibas.ch

Corresponding Author:
Florian Kuhlmey
Tel: +41 (0) 61 207 33 80
Mail: florian.kuhlmey@unibas.ch

Estimating Survival Times Using Swiss Hospital Data*

Florian Kuhlmey† and Matthias Minke†

Faculty of Business and Economics, University of Basel

June 8, 2018

Abstract

We compare and evaluate two different approaches to estimate overall survival curves from censored data of recurrent events: (1) standard survival time analysis, and (2) a multistate framework that explicitly estimates the mortality rate during censored periods. With both models, we estimate disease-specific survival curves with data from the Swiss Federal Statistical Office's medical statistics on hospitals (MedStat). Using cancer registry data as a benchmark for overall survival, we find that the accuracy of survival time estimates based on the multistate model are not superior to the simpler single-risk model. Although the computationally demanding multistate model is less accurate in predicting survival times, it may nevertheless be useful if intermediate transitions are the targeted issues.

JEL classification C41; C53; I12.

Keywords Survival analysis; multistate-model; data simulation; hospital discharge data.

1 Introduction

Mortality continues to be a widely used measure for the success (or failure) of therapies, preventive measures and even health systems as a whole. Information on cause-specific mortality rates often guides policy makers and health practitioners in their decisions on how to allocate health budgets. It is therefore important to obtain accurate estimates of mortality and survival rates. For cancer, as a leading cause of death, many countries systematically collect data on new cases, follow patients' histories, and report outcomes, which are often differentiated by socio-economic or regional characteristics. In Switzerland, cantonal and national cancer registries report prevalence, incidence, and survival for all important cancer types. For other diseases, these data are not always available, and the identification of disease-specific survival is therefore more challenging.

The statistics on the causes of death register the number of deaths per year attributed to a specific cause, but because they do not include the time of diagnosis, they do not reveal

*This paper was supported by the Swiss National Science Foundation (National Research Program 67 "End of life", grant number: NWW1513), and by the WWZ-Forum (FV-39). We thank Mehdi Farsi, Beat Hintermann, Ueli Matter, and Klazien Matter-Walstra for helpful comments and suggestions.

†Address: Peter-Merian-Weg 6, CH-4002 Basel, Switzerland.

disease-specific survival times, i.e., the time between diagnosis and death. Clinical randomized controlled trials (RCTs) are essential to determine the (causal) effect of a distinct therapy and are therefore the basis for approval decisions, but are less appropriate to assess the overall performance of treatments, since they typically do not reflect the average treatment success for the patient population and for the actual mixture of treatments. Unfortunately, a comprehensive data set that covers a broad set of diseases does not exist in Switzerland, and the same is true for many other OECD countries.

We propose to use hospital discharge data to deliver comparable survival curves for different diseases. The advantage of hospital data is that it is routinely collected for administrative purposes and is therefore an inexpensive way for researchers to obtain mortality data. The disadvantage is that it only covers events which occur during a patient's hospitalization episodes. This implies that the patient-specific observation time will start at the first (in-hospital) diagnosis and cease at the end of the last in-hospital spell with either a transition to death or censoring. Hospital records naturally only keep track of in-hospital spells, and corresponding patient histories therefore only contain the in-hospital deaths. A linkage of hospital data with death registries would be a way to overcome this problem, but for privacy reasons, at least in the Swiss case, this seems infeasible.[1]

This leaves two options to estimate survival times given this kind of data. The first is to use the in-hospital mortality as a proxy for overall mortality.[2] Technically, this method requires that one observation per patient is created which uses the date of the patient's first hospitalization with a given diagnosis as the onset of risk and the date of the patient's last discharge as the event time (in the case of death) or the censoring time (if the patient is discharged). We label this the 'single-risk model', since the only transition that we model is the patient's death at some point. The second option is to exploit the complete record of in- and out-of-hospital spells, where the latter follow as the complementary set of the hospital spells. This allows us to build a multistate model in which patients can either enter the next (out-of-) hospital spell or the state of death. We label this model the 'multistate model'.

The standard single-risk model only is convincing, if there is no selection bias owing to patients who (do not) die in the hospital. This requires that censoring occurs randomly – an assumption that does not seem plausible ex ante, since the probability of observing a patient in the hospital may systematically depend on his or her state of health. It is not even clear a priori in which direction the systematic censoring would bias the mortality approximation that is estimated based on in-hospital mortality rates. If patients with a bad health state tend to die out of hospital (e.g., in a hospice or at home), their probability of being readmitted to hospital is smaller than for patients with a relatively better health state, and we will underestimate the

[1] A study that links discharge data with the cancer registry is Ayanian et al. (1993), who analyze (with US data) differences in the stage of the disease at detection and stage-specific survival in breast cancer patients with varying insurance coverages.

[2] Bucholz et al. (2016), for example, investigate differences in long-term outcomes of patients with acute myocardial infarction (AMI) for Californian data. They find that long-term survival variations between patients from high-performing hospitals and low-performing hospitals stem from differences in the 30-day mortality that persist over time.

true mortality rate. If, on the other hand, a worsening health state increases the probability of being readmitted, we would overestimate the true mortality rate.

One solution would be to control for patient-specific factors (over time) which drive the health state. In practice, however, a significant share of patients' heterogeneity remains unobserved. Few existing models allow for potentially *informative censoring*. Among them, e.g., Danieli et al. (2012) evaluate different methods to cope with systematic censoring using simulated data to estimate net survival, relative to a baseline survival; Scharfstein & Robins (2002) use prognostic factors and allow for unmeasured factors to derive unbiased survival rates; and Wang et al. (2001) allow for informative right-censoring in the presence of recurrent events by assuming a latent variable model. The latent variable multiplicatively shifts the hazard rate, i.e., adds frailty to the transition structure.

We propose the aforementioned multistate model as an alternative method to the standard single-risk time-to-event approach, where we estimate the unobserved out-of-hospital behavior using the information on patients that are being readmitted. In the sense that we explicitly estimate the in- and out-of-hospital transition rates by the sequence number of the spell in every patient's history, we account for potentially informative censoring after a patient's last hospital discharge, without requiring assumptions on the latent variable. Note that each competing risks experiment considered by the multistate model assumes random censoring. What allows our model to cope with (potential) informative censoring is the stratification according to the sequence of in- and out-of-hospital spells, which means we allow for different in- and out-of-hospital transition rates, that vary over the sequence of admissions.

In the following, we first fit a standard single-risk model to hospital discharge data and predict survival times for a selection of diseases. In a second step, we develop a multistate model and therefore consider a more complex setup to include and estimate the out-of-hospital mortality. Section 2 introduces both models, which we fit in Section 4 to the data described in Section 3. Finally, we contrast the estimates of both models with available observational data from the National Institute for Cancer Epidemiology and Registration (NICER) in Section 5. Section 6 concludes.

2 Methodology

In this chapter, we present two alternative approaches to estimate survival time from incomplete data. Incomplete in our context means that we (need to) allow for right-censoring of observations. The first approach, presented in Section 2.1, is a standard single risk survival model with piecewise constant hazard rates. We use the thus estimated hazard function to predict disease-specific survival curves. The second is an approach that we propose in order to relax the assumption of random censoring inherent in the standard model. It consists of two parts. In Section 2.2, we present an adaption of the competing risks estimation for unobserved out-of-hospital mortality from Farsi & Ridder (2006). In Section 2.3, these results are embedded into a multistate framework and combined with the simulation algorithm from

Blaser et al. (2015) in order to finally yield disease-specific survival curves.

Although these models can be applied to any survival time data with (potentially) multiple recurrent observations, we illustrate them for the case of hospital discharges, which implies the following: The same patient might be observed several times, i.e., over the course of many hospitalizations. At the end of every hospitalization, we observe whether he or she died or was discharged, in which case an out-of-hospital spell starts that lasts until the next hospitalization or the end of our observation time. The point in time in which the patient dies in his or her last out-of-hospital spell, remains unknown.

2.1 The standard single-risk approach

For the standard single risk approach, we regard the initial hospitalization of a patient as the start of being at risk, and the end of the last observable in-hospital spell as the end-point of our analysis. Any transitions between the first and the last hospitalization are discarded. An illustration of how patient histories are used for this single-risk model, and how this differs from the use in the multistate model is given in Figure 1 on page 7. If there is only one hospital spell, the time at risk is the length of this specific hospital spell. There is only one transition possible in this setup: To move from being alive to being dead. This is, we model death as a single-risk process. If a patient's last observable hospitalization ends with death, we observe the event of interest, whereas if a patient's last observable hospitalization ends with a discharge, the patient's outcome is considered to be censored after that point in time.

To make the single-risk estimation comparable to the multistate framework, outlined later, we work with a flexible specification and allow for a hazard function that is constant within each of the $i = 1, \ldots, I$ time intervals. This implies that we assume that event times are exponentially distributed over the given time intervals. We further include a set of patient-specific time-invariant covariates.

Let t denote the time that has passed since the onset of the risk. Formally, if t is in the ith interval, the hazard function $\mu(t)$ takes the following form:

$$\mu(t) = \exp(\boldsymbol{X}\boldsymbol{\beta})\mu_i \quad \forall \; t_{i-1} < t \leq t_i, \tag{1}$$

where μ_i is the baseline hazard rate during the ith interval. The vector $\boldsymbol{X}$ consists of all time-invariant and patient-specific covariates (such as biographical information, time- and year-dummies, or information on the treatment), and $\boldsymbol{\beta}$ is the coefficient vector. Each μ_i and $\boldsymbol{\beta}$ are to be estimated.

Interval borders are given by t_{i-1} and t_i. They indicate the start and end of the time interval i, respectively, such that $t_i - t_{i-1}$ describes its length.[3] The initial admission to a hospital, i.e., the start of the first observation, marks the onset of the risk for every patient such that $t_0 = 0$. For t in the jth time interval (i.e., for $t_{j-1} < t < t_j$) and piecewise constant hazards within

[3]The time t_{i-1} is the ending time of the interval $i-1$ and t_i the starting time of the interval $i+1$.

each of these intervals, the survival function reads:

$$S(t) = \exp\left(-\exp\left(\boldsymbol{X}\boldsymbol{\beta}\right)\left(\sum_{i=1}^{j-1}\mu_i(t_i - t_{i-1}) + \mu_j(t - t_{j-1})\right)\right). \tag{2}$$

This equation gives the survival probability until time t of a patient with individual characteristics $\boldsymbol{X}$ and translates hazard rates into a probability, accounting for the effect of covariates and allowing for piecewise constant hazard rates over a set of intervals. (For a more general treatise of the connection between hazard rates and probabilities, also in the context of a multistate model, see, e.g., Beyersmann et al. 2012, Section 2.1.)

To estimate the hazard rates and parameters of the covariates, we maximize a log-likelihood function based on the probability density functions (pdfs) of all observed transitions and on information of patients who survived until the censoring date. The contribution of a patient who is observed for the length of t to the log-likelihood function therefore reads as

$$\ell(t) = \begin{cases} f(t) & \text{if death in hospital} \\ S(t) & \text{if censored,} \end{cases} \tag{3}$$

where $f(t) \equiv \mu(t)S(t)$ is the pdf of a patient who dies at t, with $\mu(t)$ being the hazard function from (1) and $S(t)$ the survival function from (2). If a transition occurs (i.e., if the patient dies at t), the pdf is defined as the product of $\mu(t)$ with the probability that no transition occured before t. Concerning the second case, let T be the patient-specific maximum time at risk (before being right-censored). If the patient is censored (i.e., if $t > T$), then only the information that he or she has not died until then is contributed to the log likelihood function. The log likelihood function follows as

$$\log \mathcal{L}(\boldsymbol{\beta}, \mu_1, \ldots, \mu_I) = \sum_{n=1}^{N} \ell(t^n), \tag{4}$$

where N is the number of observations and t^n indicates the length of the nth observation (patient), i.e., the time at risk.[4] The contribution of each patient case depends on the event type at the end of the observation period (death or censoring) as defined in (3).

After fitting the above model, the survival curve $S(t)$ can be derived by subsituting the estimated values into (2).

2.2 The multistate approach: Obtaining the hazard rates

Farsi & Ridder (2006) exploit information on rehospitalizations to formulate a maximum like-

[4]We are inexact with indexing and have only now introduced the patient-specific index $n = 1, \ldots, N$ in the definition of the likelihood function, whereas we left it out above (e.g. in the definition of equation 3). We did this in order to accentuate what is important at each step: When defining(3) it was made clear that this index refers to a specific patient's contribution to the likelihood function. However, in this respect, the summation of all contributions is crucial, which is why we have subsequently introduced the n index. This approach carries over to the remainder of the paper.

lihood estimator of the unobserved out-of-hospital mortality by modeling two independent sequential competing risks processes: (1) when initially hospitalized patients might die or be discharged alive, and (2) when discharged they might die outside the hospital or be rehospitalized. For both sets of competing risks, the authors formulate likelihood functions for all four events and consequently estimate four hazard rate functions for Californian hospital data on the treatment of acute myocardial infarction (AMI). The aim of their analysis is to account not only for the in-hospital mortality as a quality indicator, but also for the unobserved out-of-hospital mortality. Their concern is that studies often use the in-hospital mortality rates or the probability of death within some (short) time span after discharge as a quality indicator or outcome variable of interest.[5] The problem, however, is that hospitals can to some extent influence the observed in-hospital mortality through their discharge policy, and that in-hospital mortality and post-discharge mortality is a function of discharge policies. To see this, consider two hospitals. It is possible that both deliver equal quality but differ as to their observed in-hospital mortality, because one of the two discharges faster and thus the time span during which death would be observed is shorter. A meaningful quality indicator sould therefore incorporate the effect of the discharge policy, which is the motivation of the study by Farsi & Ridder (2006).[6]

We extend their approach for multiple in- and out-of-hospital spells in order to accommodate chronic or progressive diseases. Technically, we rewrite the likelihood function such that it reflects a stratification of the data according to the sequence number of the in- and out-of-hospital spells. That is, we estimate a sequence of hazard rates for the spell-specific discharge, in-hospital mortality, readmission, and out-of-hospital mortality. In a second step, these competing risks processes are then combined into a multistate model of sequential in- and out-of-hospital episodes, and the absorbing state of being dead. We now have explicit estimates for the out-of-hospital transitions, but still cannot easily derive overall survival rates that comply with them – for the simple reason that the original data naturally lacks these transitions. We therefore simulate a cohort of patients (in the next section) who move through the multistate model in accordance with the estimated transition rates. This simulation, finally, reveals the estimated overall survival times.

The hospital discharge data used in this paper allows us to follow patients over the course of many years and therefore to generate patient histories, as illustrated in Figure 1 for two different patients. There are three kinds of states in which a patient may be: hospitalized (IH), out of hospital (OH), or dead (D). Let t measure the time spent in a given state, and let τ be the aggregate time passed since the first observation. That is, t is reset to 0 once a transition occurs, and τ is the sum of all patient-specific spell lengths. The current spell in or out of the hospital is indexed with $l = 1, 2 \ldots, L$, where L is the number of strata considered.

[5] Recent examples of such applications are found in Rahman et al. (2016) who use the 30-days-post-discharge mortality as outcome, in Martini et al. (2014) who construct the outcome using the total of in-hospital and 30 days post-discharge mortality rates, in Evans & Kim (2006) where the in-hospital mortality and 7-14-day overall mortality serve as outcome variables, and in Daysal (2012) who uses the in-hospital mortality of heart attack patients.

[6] A paper that discusses the (non-)independence of mortality and readmission rates is Laudicella et al. (2013).

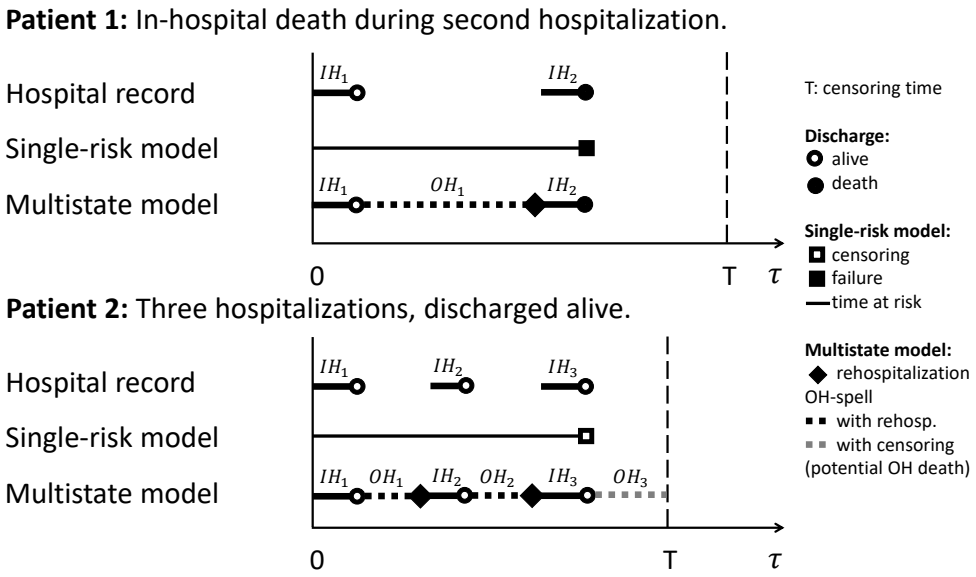
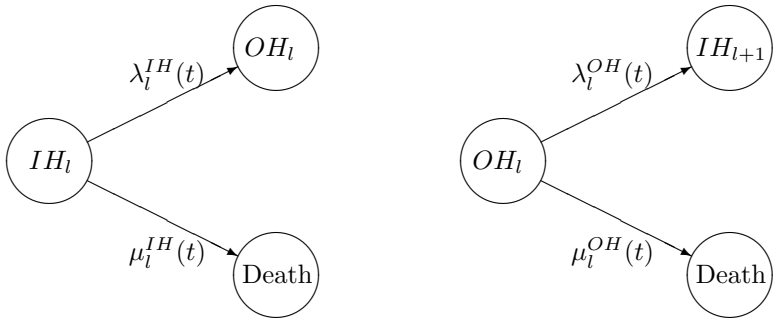


Figure 1: Illustration of patient histories.



(a) Transitions from in-hospital. *(b) Transitions from out-of-hospital.*

Figure 2: Competing risks processes for the lth in- and out-of-hospital spells.

It is assumed that a patient's history starts with the initial hospitalization with a specific diagnosis, i.e., IH_1 must necessarily occur before OH_1.

To illustrate the basic idea on how we can exploit the complete history of each patient to estimate survival curves, consider Figure 2. During the lth in-hospital spell, IH_l, a patient dies with a rate $\mu_l^{IH}(t)$ or is discharged with rate $\lambda_l^{IH}(t)$ if he or she had spent time t in that state (see panel *(a)*). Death and discharge are competing risks for exiting the hospital spell. If discharged alive, a patient enters the lth out-of-hospital spell, OH_l. An out-of-hospital spell ends with either rehospitalization or death (see panel *(b)*). Note again that time t is reset to zero after each transition. Therefore, the hazard rates in panel *(b)* describe the hazards after having spent time t in that state.

Unlike the single-risk setup used in the previous section, in the multistate model a transition to death, after a given elapse of time after the first hospitalization, might occur in any one of the $2L$ non-absorbing states. We will keep Figure 2 in mind as a reference for the following formal outline, where we develop an extended version of Farsi & Ridder (2006). For further details and derivations regarding the identification of the unobserved out-of-hospital mortality, please refer to their original paper.

In-hospital spells

To estimate hazard rates for exits from the in-hospital spells, we assume piecewise constant hazard rates over $i = 1, \ldots, I^{IH}$ intervals. This means that each of the L in-hospital spells is divided into I time intervals. We estimate hazard rates for $I \times L$ intervals and spells, each indicated by the subscript pair i, l. If a patient is not censored, all transitions are observed such that the hazard functions for death and discharge at time t read as

$$\mu_l^{IH}(t) = \exp(\boldsymbol{X}\boldsymbol{\beta})\mu_{i,l}^{IH} \quad \forall \ t_{i-1} < t \leq t_i \text{ and} \tag{5}$$

$$\lambda_l^{IH}(t) = \exp(\boldsymbol{X}\boldsymbol{\gamma})\lambda_{i,l}^{IH} \quad \forall \ t_{i-1} < t \leq t_i, \tag{6}$$

respectively. $\boldsymbol{X}$ is the same vector of time-invariant and patient-specific covariates as used in (1) and $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are the respective coefficient vectors. The interval- and spell-specific constant hazard rates are $\mu_{i,l}^{IH}$ and $\lambda_{i,l}^{IH}$. We assume that both transition processes share the same interval lengths. Interval borders are given by t_{i-1} and t_i for all i, with $t_0 = 0$.

The contribution of a patient exiting his or her lth hospital spell during the jth time interval to the likelihood function can be derived from the probability density function

$$f_l^{IH}(t) = S_l^{IH}(t) \cdot \begin{cases} \mu_l^{IH}(t) & \text{if death} \\ \lambda_l^{IH}(t) & \text{if discharged,} \end{cases} \tag{7}$$

where $S_l^{IH}(t) \equiv \exp\left(-\sum_{i=1}^{j-1}(t_i - t_{i-1})\kappa_{i,l} - (t - t_{j-1})\kappa_{j,l}\right)$ is the probability that the lth spell does not end before t, with $\kappa_{i,l} \equiv \exp(\boldsymbol{X}\boldsymbol{\beta})\mu_{i,l}^{IH} + \exp(\boldsymbol{X}\boldsymbol{\gamma})\lambda_{i,l}^{IH}$ being the hazard to exit the lth hospital spell (to either discharge or death) within the ith interval, and the sum over both of the competing risks reflects the cumulative incidence of no exit until t. Hence, the probability density function is equal to the hazard rate of the corresponding observed event multiplied by the probability of no event before t. Note that this is conceptually the same as the single-risk setup.[7] The log-likelihood for $n = 1, \ldots, N_l^{IH}$ patients entering the lth in-hospital spell is then given by:

$$\log \mathcal{L}_l^{IH}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \mu_{1,l}^{IH}, \ldots, \mu_{I,l}^{IH}, \lambda_{1,l}^{IH}, \ldots, \lambda_{I,l}^{IH}) = \sum_{n=1}^{N_l^{IH}} \log f_{l,n}^{IH}(t_l^n), \tag{8}$$

where t_l^n describes the failure time of patient n during his or her lth in-hospital spell, and $f_{l,n}^{IH}(t^n)$ is the probability density function (7) that considers the patient-specific failure event

[7] In an analogy to the single-risk setup, we can consider right-censored events. Consider a patient who is either censored or observed to transit at time t. The contribution of this patient to the log-likelihood function reads as $\ell^{IH}(t) = \left[S_l^{IH}(t)\right]^C \left[f_l^{IH}(t)\right]^{1-C}$, where C takes the value of 1 if a patient is censored (and zero otherwise). $S_l^{IH}(t)$ is the probability of observing no transition until t. If t is the censoring time, this is the contribution of such an observation to the log-likelihood. The last part, $f_l^{IH}(t)$, is given by (7) as the contribution to the log-likelihood function of a patient failing at t in his or her lth IH-spell, if not censored. For our application, however, censoring is not a problematic issue for in-hospital spells, since we observe the end of the hospital spell for almost all spells. For out-of-hospital spells, censoring is a likely event, since every spell that ends with death is observed as being censored; see below for more details.

at time t^n in spell l of patient n. The joint log-likelihood function for all patients in any of the L in-hospital spells is then given by:

$$\log \mathcal{L}^{IH}\left(\boldsymbol{\beta}, \boldsymbol{\gamma}, \left\{\mu_{i,l}^{IH}\right\}_{\forall i,l}, \left\{\lambda_{i,l}^{IH}\right\}_{\forall i,l}\right) = \sum_{l=1}^{L}\sum_{n=1}^{N_l^{IH}} \log f_{l,n}^{IH}(t_l^n). \tag{9}$$

We use curly brackets to indicate a list of all hazard rates that exist for the indicated subindices. Equation (9) will later be the basis for the estimation of spell- and interval-specific hazard rates and the covariate-related set of coefficients $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$.

Out-of-hospital spells

We wish to identify the reason why and the time when subjects leave an out-of-hospital spell. More precisely, we are interested in two competing risks that can end such a spell: Rehospitalization or death. If patients' courses were completely observable beyond their discharge from hospital, we would observe out-of-hospital deaths and could therefore estimate the readmission and out-of-hospital mortality rate in analogy to the in-hospital exit rates. With hospital discharge data, however, a researcher can (1) only observe hospitalizations and (2) only those that occur within the observed time frame $[0, T_l]$, whereas the spell-specific censoring time T_l varies at the patient level. Deaths – the second of the competing risks that we aim to model – and hospitalizations after T_l are not observed. Both events are observed as right-censored.

To estimate the hazard rates of both competing risks with incomplete data, we need to adjust the likelihood function accordingly. Farsi & Ridder (2006) show that the out-of-hospital mortality rate can – over a discrete time interval – be identified from the probability of readmission and the probability of being censored. They limit their analysis to one out-of-hospital spell (i.e., they do not stratify patients over several hospitalizations) and analyze the regression coefficients of such estimations. We extend their method in two directions. We first present a stratified version of their model before describing in the next section how the estimation results of the multiple competing risks setups can be merged to generate the overall survival rate.

Every patient who is discharged alive from a hospital spell enters an out-of-hospital spell. This implies that we need to stratify the out-of-hospital spells to L different strata as in the in-hospital case, indexed by $l = 1, \ldots, L$. Further, we assume that the hazard rates are piecewise constant over I^{OH} intervals, indexed by $i = 1, \ldots, I^{OH}$ such that the t_is define the out-of-hospital interval lengths. The number of strata in the out-of-hospital case have to be matched with the in-hospital case (to being later able to usefully combine them to one multistate model). The number of intervals that are considered in the in- and out-of-hospital data, on the other hand, might be different. The piecewise constant rates hazard functions for death outside hospital and rehospitalization at time t in interval i during the lth spell read

$$\mu_l^{OH}(t) = \exp(\boldsymbol{X}\boldsymbol{\delta})\mu_{i,l}^{OH} \quad \forall \; t_{i-1} < t \leq t_i \text{ and} \tag{10}$$

$$\lambda_l^{OH}(t) = \exp(\boldsymbol{X}\boldsymbol{\eta})\lambda_{i,l}^{OH} \quad \forall \; t_{i-1} < t \leq t_i, \tag{11}$$

respectively. Both are estimated in the maximum likelihood function below. Depending on whether a patient is observed as being censored or not, he or she contributes differently to the log-likelihood function.

If a patient is rehospitalized, his or her contribution is given by the pdf at the time of the transition. As formulated above, T_l is the patient- and spell-specific censoring time. It measures the time from the last observed hospital discharge until the end of the observable time frame in the data; i.e., it measures the time span in which a rehospitalization may potentially be observed. Whenever we observe that a patient's spell ends before T_l, i.e. if $t \leq T_l$, the patient gets rehospitalized. The probability density function of observing a hospitalization at $t \leq T_l$ given that no hospitalization or death has already occurred is

$$f_l^{OH}(t \mid t \leq T_l) = S_l^{OH}(t)\lambda_l^{OH}(t), \tag{12}$$

where $S_l^{OH}(t) \equiv \exp\left(-\sum_{i=1}^{j-1}(t_i - t_{i-1})\kappa_{i,l} - (t - t_{j-1})\kappa_{j,l}\right)$ with now $\kappa_{i,l} \equiv \exp(\boldsymbol{X\delta})\mu_{i,l}^{OH} + \exp(\boldsymbol{X\eta})\lambda_{i,l}^{OH}$. Similar to the in-hospital spell, $\kappa_{i,l}$ describes the total hazard to exit the lth out-of-hospital spell during the ith interval – due to either hospitalization or death, while only the former can be observed.

If a patient is not rehospitalized, we use the Farsi & Ridder method to identify the out-of-hospital mortality rate. The idea is to write the probability of *no* hospitalization until T_l in two parts.[8] In our version that has been extended for multiple spells, the probability of a patient being censored in his or her lth spell during the jth interval can be written as

$$\begin{aligned}\Pr(t > T_l) = & S_l^{OH}(t) \\ & + \sum_{i=1}^{j-1} \frac{\exp(\boldsymbol{X\delta})\mu_{i,l}^{OH}}{\kappa_{i,l}} \left[\exp(-\kappa_{i,l}t_{i-1}) - \exp(-\kappa_{i,l}t_i)\right] \\ & + \frac{\exp(\boldsymbol{X\delta})\mu_{j,l}^{OH}}{\kappa_{j,l}} \left[\exp(-\kappa_{j,l}t_{j-1}) - \exp(-\kappa_{j,l}t)\right].\end{aligned} \tag{13}$$

The first line on the right-hand side of (13) is the probability that a patient has not been re-admitted to hospital until t *and* is still alive ($S_l^{OH}(t)$ as defined above). This means that no transition from the current out-of-hospital spell until t occurred. The second and third lines yield the probability that a patient died out of hospital until t. It is the integral of the probability density function of dying between zero and t; here in a dissolved form for piecewise constant hazards over j intervals until t. The sum of both probabilities – i.e., of not having ended the out-of-hospital spell and of not having died until t – yields the probability that no hospitalization occurred until t. Note that the events underlying both probabilities are not observed. But the sum of these probabilities can be estimated from the data: It is the probability of being censored at t.

[8] The necessary identifying assumptions for this approach to work are (1) that no other hospitals outside our data exist, i.e. if a patient is rehospitalized, we will see him, and (2) hazard rates are assumed to be piecewise constant.

We therefore have identified two sets of equations for two sets of unknown (piecewise constant) hazard rates. Both are estimated simultanously in one likelihood function: If a patient is hospitalized, the contribution is given by equation (12); if a patient is not hospitalized, he or she contributes (13), such that

$$\ell_{l,n}^{OH}(t_l^n) = \begin{cases} \Pr(t_l^n > T_l^n) & \text{if censored} \\ f_l^{OH}(t_l^n \mid t_l^n \leq T_l^n) & \text{if hospitalized} \end{cases} \tag{14}$$

is the contribution of patient n, where $n = 1, \ldots, N_l^{OH}$ and N_l^{OH} is the number of patients entering the lth out-of-hospital spell.[9] We use the structural forms for the two sets of equations to formulate a spell-specific likelihood function which allows us to estimate the out-of-hospital mortality even though these transitions are not contained in the data:

$$\log \mathcal{L}_l^{OH}(\boldsymbol{\delta}, \boldsymbol{\eta}, \mu_{1,l}^{OH}, \ldots, \mu_{I,l}^{OH}, \lambda_{1,l}^{OH}, \ldots, \lambda_{I,l}^{OH}) = \sum_{n=1}^{N_l^{OH}} \log \ell_{l,n}^{OH}(t_l^n). \tag{15}$$

Here, too, $\boldsymbol{\delta}$ and $\boldsymbol{\eta}$ are independent of the spell number. The joint log likelihood function for out-of-hospital events for all L spells is given by

$$\log \mathcal{L}^{OH}\left(\boldsymbol{\delta}, \boldsymbol{\eta}, \left\{\mu_{i,l}^{OH}\right\}_{\forall i,l}, \left\{\lambda_{i,l}^{OH}\right\}_{\forall i,l}\right) = \sum_{l=1}^{L} \sum_{n=1}^{N_l^{OH}} \log \ell_{l,n}^{OH}(t_l^n), \tag{16}$$

where, again, we use curly brackets to indicate a list of all hazard rates that exist for the indicated sub-indices.

2.3 The multistate approach: Obtaining the survival curve

Once we have estimated interval- and spell-specific hazard rates and the covariates' effects $\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}$, and $\boldsymbol{\eta}$, we want to translate these estimates into a unifying metric, i.e., the survival rate. In a standard single-risk framework with a hazard function $\lambda(t)$ one can simply obtain $S(t)$ as described by equation (2). This is, however, not possible in a competing risks framework as there is more than one possible failure that can occur in any point in time t.

What additionally complicates our analysis is that we allow the hazard rates for all of the L in- and out-of-hospital transitions to be different. This precludes the use of more standard multistate models (such as the illness-death model with recovery) where hazard rates are assumed to follow a (semi-)Markov process.[10] Stratification according to the sequence number of in- and out-of-hospital spells allows to relax the Markov-assumption in the sense that the

[9] Note that as in the in-hospital case we added at this point the spell- and patient-specific indices l and n to the observation time t, which we previously ignored to ease presentation.

[10] Non-Markov processes in multistate models are scarce and standard software packages are basically not available (for a review see Willekens & Putter 2014, p. 384). An exception is Meira-Machado et al. (2006) who develop history-dependent estimates of hazard rates for a three-state illness-death-model. Their approach, however, relies on the observation of all transmissions and does not consider multiple spells.

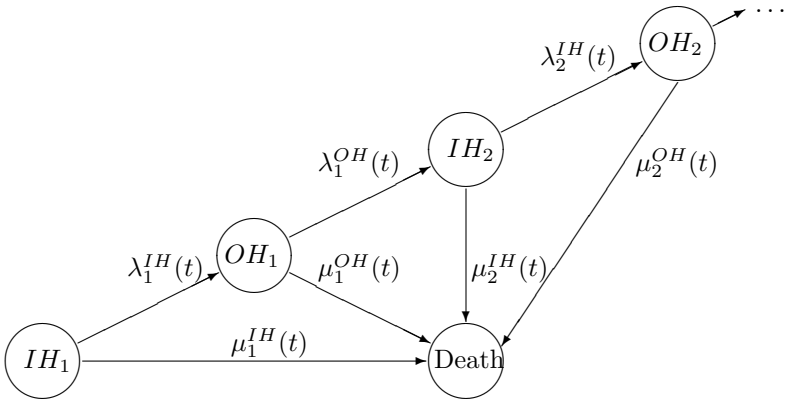


Figure 3: Multistate model.

sequence number of spells serves as proxy for characteristics such as disease progression. This implies that we allow for underlying path dependency – at least to the degree that we allow the hazard rates that apply to the, say, 5th hospital spell to be different from the hazard rates that apply to the first hospitalization.

Our multistate framework is illustrated in Figure 3. Since we defined the onset of the risk as the first hospitalization, every history starts with IH_1, which is followed by the first out-of-hospital spell OH_1 or death. For example, the spell OH_1 might end with either the second hospital spell IH_2 or death.

There are $(2L+1)$ states and $(4L-1)$ transitions that can occur at any time t: A patient can (theoretically) be in any one of the L in- or out-of-hospital spells, or he or she could be dead, which is the only absorbing state. When the patient is in any of the $2L$ transitory states, he or she can either die or move to the next transitory state. An exception is the last out-of-hospital spell for which we assume that no rehospitalization is possible, and that patients simply die in accordance with the out-of-hospital mortality rate for the last spell. This implies that we have to pick a sufficiently high L such that only as few patients as possible are affected by this inconsistency. This is also the reason why we only need to model $(4L-1)$ transitions.

To obtain the overall disease-specific survival rate, we have to rely on a simulation since the out-of-hospital deaths are not observed in our original data. We use the estimated hazard rates from the previous section and simulate a cohort of patients that moves through a multistate model as depicted in Figure 3. This yields transition times for every simulated patient and therefore reveals the in- and out-of-hospital mortality. The simulation thus allows us to translate the estimated hazard rates and the given structure of our multistate model into an overall survival rate.

To perform this Monte-Carlo-like simulation, we use a corrected version[11] of an algorithm developed by Blaser et al. (2015). An important advantage of this package is that we can

[11] While working on this study, we discovered several errors in the Blaser et al. (2015) code, which we corrected. An updated version of the R package has in the meantime been uploaded on CRAN.

incorporate time-varying and spell-specific hazard rates (i.e., we can capture the intervalization as modeled for estimation). To put the algorithm to work, we set up a patient cohort of size N_1^{IH} with cohort characteristics $\boldsymbol{X}$. Both N_1^{IH} and $\boldsymbol{X}$ are chosen to match the size of the initial patient-population and the distribution of the covariates in the data. The estimation results from the previous section are then used to set up a hazard rate matrix for all possible transitions as well as a variance-covariance matrix for any of the possible transitions of the hazard rate matrix.

The hazard rate matrix $\mathrm{H}(t)$ is a function of time t and of all estimated coefficients from the previous section: $\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\eta}, \left\{\mu_{i,l}^h\right\}_{\forall i,l,h}, \left\{\lambda_{i,l}^h\right\}_{\forall i,l,h}$, where $h \in \{IH, OH\}$. It is given by

$$\mathrm{H}(t) = \begin{array}{c} \\ IH_1 \\ OH_1 \\ IH_2 \\ \vdots \\ IH_L \\ OH_L \end{array} \begin{array}{c} \begin{array}{ccccccc} OH_1 & IH_2 & \dots & IH_L & OH_L & D \end{array} \\ \begin{pmatrix} \lambda_1^{IH}(t) & 0 & \dots & 0 & 0 & \mu_1^{IH}(t) \\ 0 & \lambda_1^{OH}(t) & \dots & 0 & 0 & \mu_1^{OH}(t) \\ 0 & 0 & \dots & 0 & 0 & \mu_2^{IH}(t) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & \lambda_L^{IH}(t) & \mu_L^{IH}(t) \\ 0 & 0 & \dots & 0 & 0 & \mu_L^{OH}(t) \end{pmatrix} \end{array}. \tag{17}$$

The rows list the states from which a transit may occur, and columns, the states into which transitions are possible. Consequently, for L in- and L out-of-hospital spells, there are $2L$ possible transitions from in-hospital states and $2L-1$ transitions from out-of-hospital. This sums to the total of $4L-1$ hazard functions, as input for the hazard matrix $\mathrm{H}(t)$. Recall that t measures the corresponding length of spell for every transitory state (i.e., t is reset to zero in every row). Put differently, H(5 days) gives the matrix of the hazard rates for all transitions after being in any of the transitory spells for 5 days.

The algorithm of this package then draws (for every simulated patient) transition times for every possible transition and constructs the corresponding transition history for any given patient. The draws of the transition times follow the timing pattern implied by the hazard rates for the respective transitions. For a translation of the individual failure events into an aggregate time frame, some additional steps are needed. Suppose there is a realized history $\{IH_1, OH_1, IH_2, OH_2, D\}$ with transition times $\{t_1^{IH}, t_1^{OH}, t_2^{IH}, t_2^{OH}\}$, implying that this patient died during his or her second out-of-hospital spell after staying for t_2^{OH} time units in that spell. We define τ as the total survival time after the initial hospitalization, such that $\tau := t_1^{IH} + t_1^{OH} + t_2^{IH} + t_2^{OH}$. By tracking τ for every simulated patient, we can count the number of deceased individuals at any given point in time and the survival rate $S(\tau)$ is then simply the number of deceased individuals over the total number of the initial patient cohort.

This approach leaves room for two sources of uncertainty: one stemming from choosing a finite cohort size, and one from parameter uncertainty, since we have to rely on estimated hazard rates and covariate-coefficients. Concerning the former, we advocate using the original number of patients. Concerning the latter, the Blaser et al. (2015) algorithm allows us to

additionally plug in a variance-covariance matrix for each of the defined transitions. For every non-zero entry of the hazard matrix (17), we therefore also fed in a variance-covariance matrix that contains on its diagonal the variance of each estimated parameter from the previous step.

3 Data

For our analysis we use two sources of data. The first is the Swiss medical statistics of hospitals (MedStat), which we use to fit both the single-risk model of Section 2.1 and the multistate model as presented in Section 2.2. The second source is the cancer registry of the Swiss National Institute for Cancer Epidemiology and Registration (NICER). It serves as an independent data source of survival times for a limited range of cancer types and as a benchmark for our survival estimates.

MedStat

We use case-level data from the Swiss medical statistics of hospitals (MedStat) which covers almost all hospitalizations in Switzerland and is maintained by the Swiss Federal Statistical Office. Although these data have been collected since 1998, we use only data from 2001 onwards, which allows us to exclude patients in our 2001-cohort who have had treatments due to the disease of interest before 2001 (see below for more details). Each case is indexed by an anonymous identifier based on the patient's given name, family name, sex, and date of birth. The encryption method claims to unambiguously assign cases to individuals in 99.7% of cases.[12] This allows us a reasonably precise construction of individuals' entire hospitalization histories between the first and last observable hospital admission during a 10-year[13] observation period. For each hospital admission, information is recorded on spell length, patients' 5-year age-group, sex, place of residence, canton of treatment, main and up to 49 secondary diagnoses (ICD-10 codes), main and secondary treatments (CHOP codes), intensive care unit (ICU) use, and the prescription of expensive drugs.[14]

For inclusion we use several criteria: We (1) consider only hospitalization between 2001 and 2014, which delivers 19,705,990 individual spells. Of this number, we exclude (2) spells of patients living outside Switzerland (2.2% of the cases), (3) newborns (5.2% of the remaining cases), (4) spells comprising mental health issues (7.3% of the remaining cases), (5) spells of

[12] See `https://www.bfs.admin.ch/bfs/de/home/statistiken/gesundheit/erhebungen/ms.assetdetail.230439.html`, last accessed June 2018

[13] The encryption yields a stable identification code over that time span (see `https://www.bfs.admin.ch/bfs/de/home/statistiken/gesundheit/erhebungen/ms.assetdetail.230439.html`, last accessed June 2018).

[14] For detailed information on all variables included in MedStat see `https://www.bfs.admin.ch/bfs/de/home/statistiken/gesundheit/erhebungen/ms.assetdetail.1922896.html`, last accessed June 2018. For information on the International Statistical Classification of Diseases (ICD) and the ICD-10 classification for diagnoses, see `http://apps.who.int/classifications/icd10/browse/2014/en`, last accessed June 2018. For information on CHOP codes, which classify (surgical) treatments in Switzerland, see `https://www.bfs.admin.ch/bfs/de/home/statistiken/gesundheit/gesundheitswesen.assetdetail.483959.html`, last accessed June 2018.

Table 1: Summary statistics of MedStat 2001-2014.

Observations		15,079,338
Patients		5,631,757
In-hospital deaths		330,181
	Mean (Std. Dev.)	*Median*
Duration in hospital (days)	7.573 (13.064)	4
Spells per patient	2.672 (2.971)	2
		Share in %
Rehospitalized		54.1
Patients dead		5.9
Female		55.7
Swiss nationality		84.2
Disease category		
C: Cancer		6.7
G: Nervous system		3.1
I: Circulatory system		12.7
J: Respiratory system		4.6
K: Digestive system		9.1
S: Injuries		9.5
Comorbidities		
0		32.5
1		15.6
2		12.4
3		9.6
4		7.5
5 or more		16.7
Age		
20 to 30		9.1
30 to 40		12.7
40 to 50		11.8
50 to 60		14.5
60 to 70		17.1
70 to 80		18.5
80 to 90		13.6
90 to 100		2.6

patients without nationality, or with the ID number "0" (0.3% of the remaining cases), (6) patients below the age of 20 (9.2% of the remaining cases), and finally (7) cases marked as double entry due to year-change during the spell or with a length of stay of zero days or longer than two years (1.7% of the remaining cases). This leaves 15,079,338 observations.

Descriptive statistics of the hospital spells that comply with our inclusion criteria are shown in Table 1. In our analysis, we focus on cancer and diseases of the digestive system, defined as ICD-10 categories C and K, respectively, which represent 6.7% and 9.1% of the main diagnoses. Our selection of diseases out of these two categories is displayed in Table 2. Diseases are identified by ICD-10 codes of main and secondary diagnoses.[15] We are ultimately interested in the

[15]A comment on the coding principles of the ICD system: As mentioned above, ICD – the International Statistical Classification of Diseases – is a hierarchical system that describes a disease using a combination of a leading letter that is followed by 2 to 5 numbers. In our selection and definition of diseases, we use different levels of aggregation. Take the example of a lung cancer located at the main bronchus. The corresponding

average survival times of patients admitted with a specific disease after initial hospitalization due to that disease. Typically, a patient's hospitalization history does not consist of only one, but many diagnoses. We therefore face the problem of how to assign a patient to a specific disease. We chose the following procedure: We search every case's principal and first 10 of the 49 reported secondary diagnoses for a specific disease. If we find the diagnosis, we search for subsequent spells of the respective patient, which allows us to construct the patient's history. Hospitalizations before this initial diagnosis-specific spell are disregarded. This implies that if a patient has more than one of the diseases of interest, he or she will be included in each of the disease-specific patient cohorts.[16]

For the multistate model, we construct out-of-hospital spells based on discharge dates and subsequent readmission dates. If a readmission does not occur within the observed time span, the out-of-hospital spell is censored and this case contributes to the hazard rate estimation that no rehospitalization occurred until this point in time. The data on the in-hospital spells is used to estimate the in-hospital mortality and discharge rate, and the data on the out-of-hospital spells is the input to estimate hazard rates of rehospitalization and out-of-hospital deaths. All are then combined in the simulations to determine the disease-specific survival curves reported below.

The out-of-hospital hazard rate estimates are found to be sensitive to the time horizon over which patients are (potentially) observable. Our approach to account for this problem is to work with patient cohorts for which the observable time horizon is equally long. The relative cohort sizes can be deduced from Table 2. This ensures that, in principle, the probability of being observed once again after discharge (i.e. of being rehospitalized) is equally distributed. A cohort consists of all patients who are initially hospitalized in a given year and are subsequently followed up during a given time span. There is a trade-off between the length of follow-up and the number of cohorts that we consider. Since we are interested in the first diagnosis of a given disease, we use the first three years (1998-2000) to exclude patients from later cohorts who were already treated for this disease before.[17] Since the follow-up period must be of equal length for all cohorts, and in order to have a reasonable follow-up period (to make statements about medium- to long-term survival), we choose a follow-up period of 7 years. This implies that we consider 7 cohorts and include patients initially treated in the years 2001 to 2008. The first cohort is followed until the end of 2007, and the last until the end of 2014. This cohort

ICD-10 code is C34.0. The "C" indicates a malignant neoplasm, the first two numbers indicate that this neoplasm is located at the bronchus or lung, and the ".0" specifies the precise location at the main bronchus. For cancer, we define diseases by the first two digits. But we also use a higher aggregation level: diagnoses beginning with K7 are used to indicate (chronic) liver diseases—and consequently comprise a broader range of diseases (e.g. K70 alcoholic liver disease, K71 toxic liver disease, K73 chronic hepatitis).

[16] An alternative approach would be to define a hierarchy of diagnoses and assign every patient's history to only one disease according to this hierarchy. This would ensure that every patient is only used once. Every patient would be attributed to only one condition and we would ignore potential other diseases of interest that this patient might have. However, we are not aware of a natural hierarchy of diseases that could be applied here.

[17] Note that during the initial years of record-keeping, not all hospitals participated. For the analysis of chronic diseases (see below), the data for the first cohorts may be biased, because they contain some individuals who are not 'true' newly diagnosed, but who where diagnosed before and not reported.

Table 2: Summary statistics for the selection of chronic diseases (first diagnosed 2001 to 2008, follow-up 7 years).

ICD-10	C16	C18	C20	C22	C25	C34	C50	C61	C92	K70-K77
						(in '000)				
Observations	29.8	96.6	45.8	24.9	29.4	128.9	197.6	179.5	15.5	303.7
Patients	7.5	24.0	9.9	7.0	9.6	32.9	51.8	47.1	3.9	72.6
In-hospital deaths	3.3	7.6	2.9	4.1	5.1	17.5	8.5	10.2	2.0	20.0
						(in %)				
In-hospital deaths	44.0	31.7	29.9	58.8	53.1	53.2	16.4	21.7	51.2	27.5
Age ≥ 65	67.1	71.1	64.6	64.2	70.5	60.8	45.6	73.3	58.3	43.0
Male	60.1	53.3	59.5	69.7	48.4	66.8	1.0	99.9	55.1	58.1
Year										
2001	13.9	13.0	13.4	11.4	11.4	12.8	13.2	11.7	12.0	10.1
2002	12.0	12.0	11.9	12.4	11.4	11.5	12.6	11.9	10.6	10.0
2003	12.4	12.5	11.8	11.0	12.0	12.1	12.7	12.1	11.8	10.5
2004	12.2	12.3	12.4	12.5	11.9	11.7	12.5	12.4	11.5	11.8
2005	11.9	12.3	13.0	13.6	12.9	13.5	12.2	12.8	14.7	14.5
2006	12.7	12.4	12.5	13.1	13.1	12.6	12.1	12.5	13.0	14.9
2007	12.2	12.8	11.9	12.9	13.2	12.5	12.1	13.3	12.8	14.6
2008	12.6	12.7	13.1	13.2	14.1	13.3	12.5	13.3	13.6	13.5
Spell length						*(in days)*				
25%	2	2	2	2	2	2	2	2	2	2
median	6	8	7	6	8	6	5	6	7	7
mean	11.4	11.8	11.7	11.0	12.9	10.9	8.5	9.7	15.1	11.4

Decoding of ICD-10 keys. C16: Malignant neoplasm of stomach, C18: Malignant neoplasm of colon, C20: Malignant neoplasm of rectum, C22: Malignant neoplasm of liver and intrahepatic bile ducts, C25: Malignant neoplasm of pancreas, C34: Malignant neoplasm of bronchus and lung, C50: Malignant neoplasm of breast, C61: Malignant neoplasm of prostate, C92: Myeloid leukaemia, and K70-K77: Diseases of liver.

Table 3: Available cancer survival data from NICER.

ICD code	Cancer site	Multistate model convergence
C16	Stomach	✓
C18*	Colon	✓
C20*	Rectum	✓
C22	Liver and intra-hepatic bile ducts	✓
C25	Pancreas	✓
C33*	Trachea	
C34*	Bronchus and lung	✓
C50	Breast	✓
C56	Ovary	
C92*	Myeloid leukaemia	✓
C94*	Other leukaemias of specified cell type	

* NICER groups survival data for some cancer types into subcategory pairs: C18/C20, C33/34, C92/C94.

approach ensures that out-of-hospital censoring remains comparable over the years.

NICER data

As a benchmark for both the single-risk and the multistate approach, we consult registry data on overall survival times for a selected range of cancer types. These data come from the Swiss National Institute for Cancer Epidemiology and Registration (NICER). NICER does not depend solely on hospital records, but collects data on newly diagnosed cancer cases and follows patients over time. The registry provides an ideal data source for comparative studies as it covers patients' whole history of patients, in particular it includes both in- and out-of-hospital deaths. It differentiates between cause-specific deaths (i.e., only deaths that can be attributed to a specific cancer) and the crude overall death rate (i.e., all-cause deaths of persons diagnosed with a specific cancer). MedStat does not contain any information on (single) causes of death other than those recorded in the diagnoses. We therefore compare the estimates based on the MedStat to the crude survival measure from the NICER data. A major shortcoming of the NICER data, however, is its limited availability for only a number of cancer types (further issues are discussed in Section 5). Table 3 lists the availability of NICER data and cancer types for which we have a multistate estimate of out-of-hospital hazard rates and therefore an estimate of survival times. The multistate estimation is not available for all cancer types listed by the NICER. The reason is that the estimation of out-of-hospital mortality hazards did not converge for all diseases, e.g. because of a too small N or not sufficient readmissions.

4 Results

4.1 Single-risk model

For the estimation of the single-risk model as developed in Section 2.1, we specify the hazard function (1) as

$$\mu(t) = \exp(\text{old } \beta^{old} + \text{male } \beta^{male} + \sum_{y=2002}^{2008} \text{cohort}^y \ \beta^y)\mu_i \quad \forall \ t_{i-1} < t \leq t_i, \tag{18}$$

where $i = 1, \ldots, 8$ indicate time intervals. We choose the time intervals such that we capture potential breaks of the hazard rate and mimic the interval structure of the multistate model. In the multistate model we work with 6 in- and out-of-hospital spells with 3 time intervals over each spell (for further details see the next section). In the single-risk process, we discard the intermediate transitions and focus on the single event of interest – a transition to being dead. This implies that there is no 1-to-1 relationship of the intervalization between the two models. The vector of covariates is given by $\boldsymbol{X} = [\text{old}, \text{male}, \text{cohort}^{2002}, \ldots, \text{cohort}^{2008}]$. The dummy variables indicate patients above the age of 65, and of male sex, and cohort^y is 1 if a patient was first diagnosed in year y. We estimate this model for all cancer types that have a tick-mark in Table 3.

The results of a log-likelihood estimation are summarized in Table 4. Each column refers to the estimation of (18) for patients initially hospitalized for the cancer category code indicated in the column header. The effect of a dummy being 1 for the hazard rate of dying is given by $\exp(\beta^x)$ for any x from the vector of covariates $\boldsymbol{X}$ such that a positive [negative] value implies a higher [lower] mortality. For all diseases, being old substantially increases mortality. Males exhibit statistically significant higher mortality rates than women for half of the cancer categories. Concerning the year-effects, recall that we considered the cohorts of 2001 to 2008. Consequently, a significantly negative sign for any of the later cohorts indicates that the mortality risk was lower when compared to the 2001 cohort. Whether the 2008 cohort has a different mortality rate than the, say, 2006 cohort cannot be deduced. We find that for some diseases there was no significant mortality reduction (C16, C22, C34) for later cohorts, whereas for others mortality decreased – at least compared to the 2001 cohort. Decreases in mortality are especially pronounced for colon cancer (C18) and breast cancer (C50). Whether this implies that treatments became much more successful over the past decade for these conditions or whether the reduction in mortality reduction is due to unobserved heterogeneity of the patient cohort (i.e., that just a different set of patients gets those diagnoses that look equal to us) cannot be said with certainty.

The second set of estimated coefficients specifies the hazard rates during each of the intervals. The seven cut-offs which determine the eight intervals are set to 1 day, 7 days, 21 days, 90 days, 365 days, 730 days, and 1,460 days. Recall that we follow cohorts for six subsequent years such that the maximum follow-up time is seven years (six years plus the rest of the year of the first diagnosis). The hazard during the first day of hospitalization is measured by the constant.

Table 4: Single risk setup: Estimation results.

	C16	C18	C20	C22	C25	C34	C50	C92
Covariates								
old	0.191 ***	0.246 ***	0.369 ***	0.242 ***	0.374 ***	0.232 ***	0.436 ***	0.944 ***
male	0.100 **	0.044	-0.016	0.046	0.099 ***	0.142 ***	0.186 *	0.051
2002	0.003	-0.050	-0.112	0.079	0.054	-0.007	-0.109 *	-0.031
2003	-0.118	-0.043	-0.095	0.109	-0.065	0.026	-0.047	0.098
2004	-0.050	-0.129 **	-0.062	0.061	-0.029	-0.025	-0.216 ***	0.028
2005	-0.037	-0.134 **	-0.219 **	-0.061	-0.071	-0.090 **	-0.260 ***	-0.195 *
2006	-0.059	-0.086	-0.078	0.084	-0.097	0.034	-0.210 ***	0.052
2007	0.027	-0.142 **	-0.138	-0.009	-0.122 *	0.011	-0.267 ***	0.019
2008	-0.115	-0.142 **	-0.237 **	0.009	-0.164 **	-0.033	-0.271 ***	-0.014
Hazard rates and shifters								
$\ln(\mu_1)$	-5.304 ***	-5.690 ***	-5.772 ***	-4.515 ***	-4.819 ***	-4.705 ***	-6.228 ***	-4.727 ***
shifter_2	-0.394 *	-0.288 *	-0.343	-0.426 ***	-0.418 ***	-0.573 ***	-0.497 ***	-0.486 ***
shifter_3	-0.340 *	-0.391 ***	-0.792 ***	-0.597 ***	-0.406 ***	-0.828 ***	-0.844 ***	-1.003 ***
shifter_4	-0.961 ***	-1.320 ***	-1.500 ***	-1.241 ***	-1.003 ***	-1.414 ***	-1.789 ***	-1.732 ***
shifter_5	-1.711 ***	-2.415 ***	-2.469 ***	-2.274 ***	-1.656 ***	-2.043 ***	-2.712 ***	-2.677 ***
shifter_6	-2.239 ***	-2.538 ***	-2.552 ***	-2.778 ***	-2.263 ***	-2.544 ***	-2.888 ***	-3.186 ***
shifter_7	-2.762 ***	-2.733 ***	-2.710 ***	-3.171 ***	-2.883 ***	-3.217 ***	-2.865 ***	-3.674 ***
shifter_8	-2.958 ***	-2.619 ***	-2.435 ***	-3.352 ***	-3.144 ***	-3.267 ***	-2.312 ***	-3.739 ***
N	35,488	126,651	53,544	30,270	40,292	146,954	269,579	17,661

Decoding of ICD-10 keys. C16: Malignant neoplasm of stomach, C18: Malignant neoplasm of colon, C20: Malignant neoplasm of rectum, C22: Malignant neoplasm of liver and intrahepatic bile ducts, C25: Malignant neoplasm of pancreas, C34: Malignant neoplasm of bronchus and lung, C50: Malignant neoplasm of breast, and C92: Myeloid leukaemia. *Covariates.* All covariates are dummy variables. They indicate if patients are 65 or older, are male, and to what cohort they belong (2001 is the baseline cohort). *Hazard rates and shifters.* μ_1 is the hazard rate of dying to die during the first interval. The hazard rate of dying during the ith interval is given by $\mu_i = \exp(\ln(\mu_1) + \text{shifter}_i)\ \forall\ i \in [2, 8]$. The seven corresponding cut-offs are at 1 day, 7 days, 21 days, 90 days, 365 days, 730 days, and 1,460 days.

For example, in the case of C16, it is given by $\mu_1 = \exp(-5.304) = .005$. For the 2nd to the 8th interval, shifter$_i$ indicates the effect on the hazard rate during the ith time interval relative to the first day of hospitalization. For a patient diagnosed with C16, the hazard rate during the second interval (i.e., for days 2 to 6) is given by $\mu_2 = \exp(-5.304 - 0.394) = 0.0034$, and accordingly for the remaining intervals. A negative sign therefore indicates a lower mortality rate during an interval relative to the first day of hospitalization. As we can see, we have the same pattern for the range of diseases, namely that the mortality rate decreases over time. This indicates that if people die, they die shortly after their first hospitalization.

We use the estimation results to predict the survival times of the patients in our sample who are initially hospitalized with the cancer types listed in Table 3. That is, we use the interval-specific hazard rates to compute the survival function $S(t)$ using its fundamental relationship with the hazard rates given in equation (2). In Section 4.4, the survival rates obtained using this approach are then compared to the multistate model estimates and the observed mortality rates from the NICER data in Figure 9 on page 32.

4.2 Multistate model: Estimation

The model from Section 2.2 is specified as follows. We set $L = 6$, i.e., we stratify the data into patients in their 1st, 2nd, 3rd, 4th, 5th, and (more than) 6th in- or out-of-hospital spell. The approach thus allows for some history-dependence in the sense that transition probabilities may be different for every spell number. The number of intervals which divide a given spell into parts is set to three, i.e., $I^{IH} = I^{OH} = 3$, which implies that two cut-off points need to be chosen for every likelihood function. The cut-offs for in-hospital spells are set at the 25 and 75 percentile of the in-hospital spell length (see Table 2), for out-of-hospital spells after two and four years.[18] We maximize two log-likelihood functions separately: Equation (9) serves to estimate the hazard rates and covariate-coefficients of the competing risks to end the in-hospital spells, and equation (16) for the out-of-hospital case. Covariates are the same as in the single-risk case discussed above. As indicated in Section 2.2, we assume the effect of the covariates to be equal for every spell, but differ between in- and out-of-hospital spells.

The hazard functions describing the rates according to which a patient dies or is discharged during the ith interval (i.e., $\forall\ t_{i-1} < t \leq t_i$) in his or her lth in-hospital spell are given by

$$\mu_l^{IH}(t) = \exp(\text{old } \beta^{old} + \text{male } \beta^{male} + \sum_{y=2002}^{2008} \text{cohort}^y\ \beta^y)\mu_{i,l}^{IH} \text{ and} \tag{19}$$

$$\lambda_l^{IH}(t) = \exp(\text{old } \gamma^{old} + \text{male } \gamma^{male} + \sum_{y=2002}^{2008} \text{cohort}^y\ \gamma^y)\lambda_{i,l}^{IH}, \tag{20}$$

respectively. The hazard rates that govern the lth out-of-hospital spell of a patient in the ith

[18] For C50, we use 500 and 1,000 days as cut-offs which ensure convergence.

interval to end by death or rehospitalization are expressed as

$$\mu_l^{OH}(t) = \exp(\text{old } \delta^{old} + \text{male } \delta^{male} + \sum_{y=2002}^{2008} \text{cohort}^y \ \delta^y)\mu_{i,l}^{OH} \text{ and} \tag{21}$$

$$\lambda_l^{OH}(t) = \exp(\text{old } \eta^{old} + \text{male } \eta^{male} + \sum_{y=2002}^{2008} \text{cohort}^y \ \eta^y)\lambda_{i,l}^{OH}, \tag{22}$$

respectively. The former two equations are specifications of (5) and (6), the latter two correspond to (10) and (11).

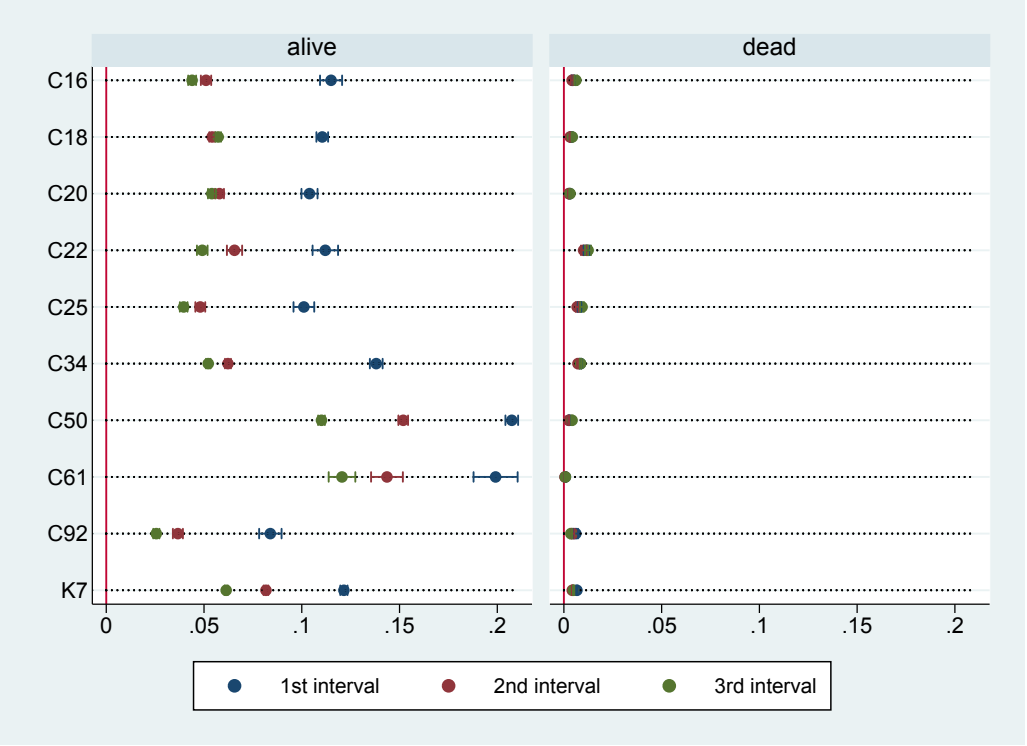
Each log-likelihood function requires to specify a total of $2 \times I \times L$ hazard rates for the competing risks, as well as $2 \times \text{length}(\boldsymbol{X})$ estimates for the coefficients of the covariates (i.e., the vectors $\boldsymbol{\beta}, \boldsymbol{\gamma}$ as well as $\boldsymbol{\delta}, \boldsymbol{\eta}$). The maximization of (9) and (16) with as many variables requires a large data set. To improve the convergence for smaller data sets, however, one can reduce the number of hazard rates that need to be estimated – at the cost of reduced flexibility. Instead of $I \times L$ hazard rates for both the IH or OH cases, it is sufficient to estimate only $I + L - 1$ hazards, if we restrict the hazards of subsequent spells to being shifted linearly for all the intervals of that spell. With $I = 3$ and $L = 6$, we need three estimates for the first spell plus five additional estimates for shifters that describe hazard rate levels in the subsequent spells.

The procedure is as follows: For any of the four possible transition types (discharge, death IH, rehospitalization, death OH), a $I \times L$ matrix of hazard rates needs to be specified. For example, the hazard matrix for being discharged alive from an in-hospital spell has the following form:

$$\Lambda^{IH} = \begin{array}{c} \\ \text{Interval} \end{array} \begin{array}{c} \text{Spell number} \\ \begin{pmatrix} \lambda_{1,1}^{IH} & \lambda_{1,2}^{IH} & \lambda_{1,3}^{IH} & \lambda_{1,4}^{IH} & \lambda_{1,5}^{IH} & \lambda_{1,6}^{IH} \\ \lambda_{2,1}^{IH} & \lambda_{2,2}^{IH} & \lambda_{2,3}^{IH} & \lambda_{2,4}^{IH} & \lambda_{2,5}^{IH} & \lambda_{2,6}^{IH} \\ \lambda_{3,1}^{IH} & \lambda_{3,2}^{IH} & \lambda_{3,3}^{IH} & \lambda_{3,4}^{IH} & \lambda_{3,5}^{IH} & \lambda_{3,6}^{IH} \end{pmatrix} \end{array} \tag{23}$$

Corresponding matrices exist also for the other three transition types: For the in-hospital mortality, the hazard matrix consists of all $\mu_{i,l}^{IH}$ and is labeled M^{IH}; for the out-of-hospital spells the matrices are labeled Λ^{OH} for rehospitalizations and M^{OH} for out-of-hospital deaths. Instead of estimating 18 different hazard rates for every hazard matrix, we only estimate hazard rates for the first spell plus one shifter for each of the five subsequent spells. This means that we estimate $\lambda_{1,1}^{IH}$, $\lambda_{2,1}^{IH}$, and $\lambda_{3,1}^{IH}$ (the three hazards for the first spell) as well as $\lambda_2^{IH}, \lambda_3^{IH}, \ldots, \lambda_6^{IH}$ (the shifters of the hazard rate for the subsequent spells). Any of the hazard rates of one of the subsequent spells is then computed by $\lambda_{i,l\geq 2}^{IH} = \lambda_{i,1}^{IH} \cdot \lambda_l^{IH}$. The same procedure is applied to specify M^{IH}. All in-hospital hazard rates are then estimated maximizing (9) – which also yields estimates for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. For the second log-likelihood function, (16), finally, we also use this method to identify Λ^{OH} and M^{OH} (along with $\boldsymbol{\delta}$ and $\boldsymbol{\eta}$). For each of the four transition types and for all diseases listed in Table 2, the estimated disease-specific hazard rates of the first spell are plotted in Figure 4.

(a) In-hospital spells: $\lambda_{i,1}^{IH}$ ('alive') and $\mu_{i,1}^{IH}$ ('dead').



(b) Out-of-hospital spells: $\lambda_{i,1}^{OH}$ ('alive') and $\mu_{i,1}^{OH}$ ('dead').

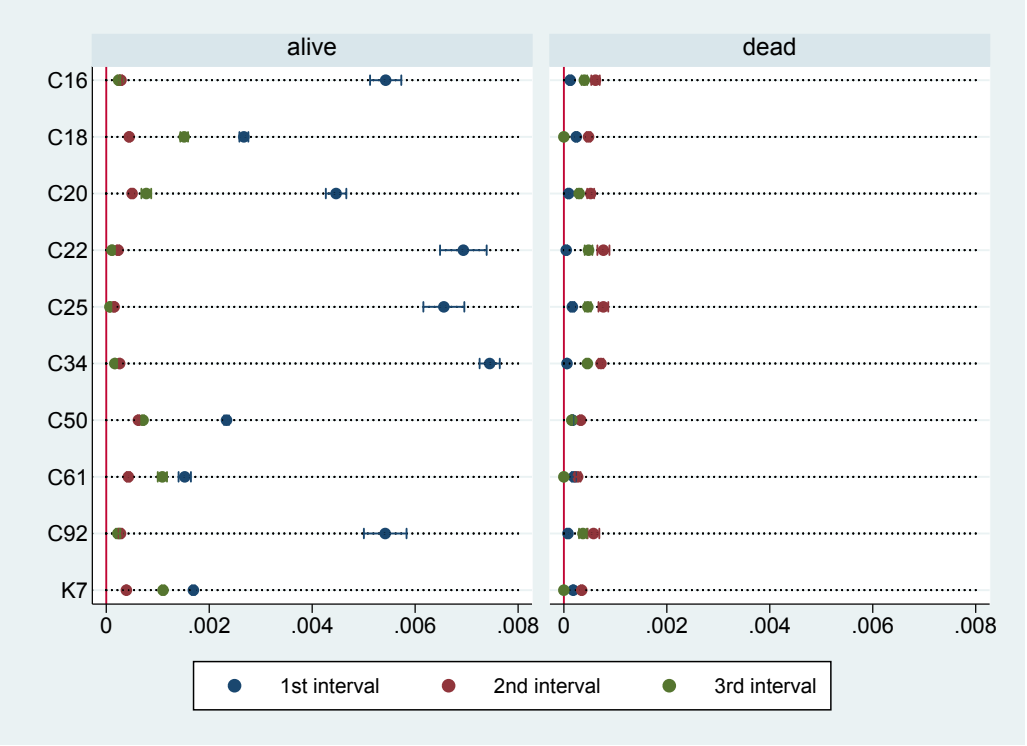
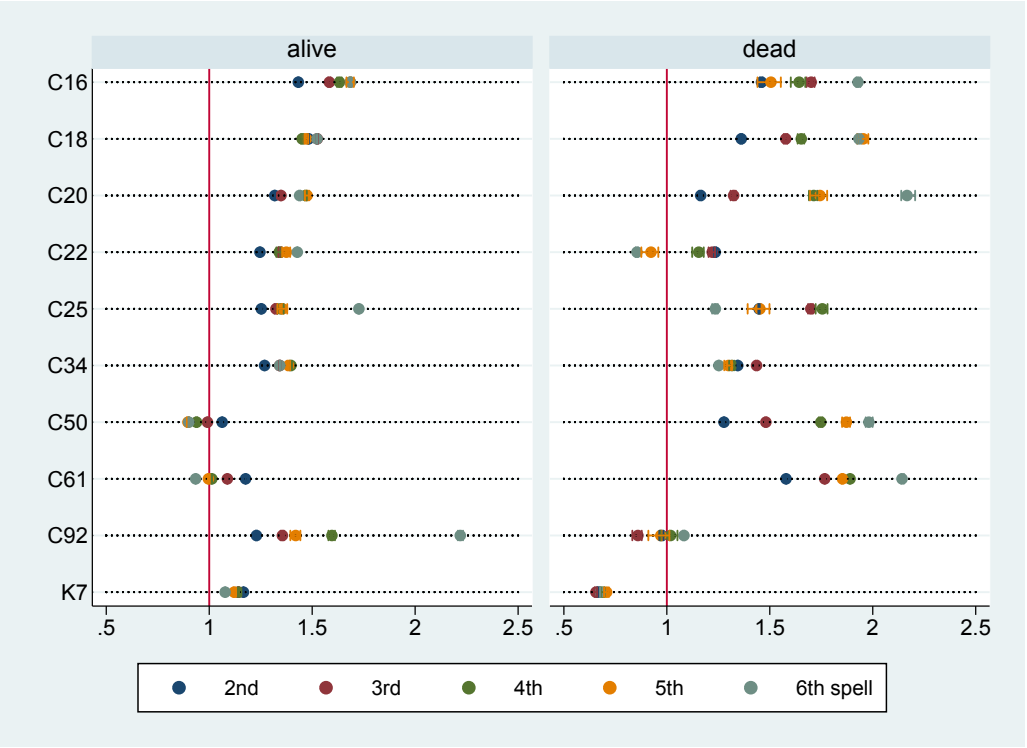


Figure 4: Hazard rate estimations for the intervals of the first spells.

Decoding of ICD-10 keys. C16: Malignant neoplasm (MN) of stomach, C18: MN of colon, C20: MN of rectum, C22: MN of liver and intrahepatic bile ducts, C25: MN of pancreas, C34: MN of bronchus and lung, C50: MN of breast, C61: MN of prostate, C92: Myeloid leukaemia, and K70-K77: Diseases of liver.

(a) In-hospital spells: λ_l^{IH} ('alive') and μ_l^{IH} ('dead').



(b) Out-of-hospital spells: λ_l^{OH} ('alive') and μ_l^{OH} ('dead').

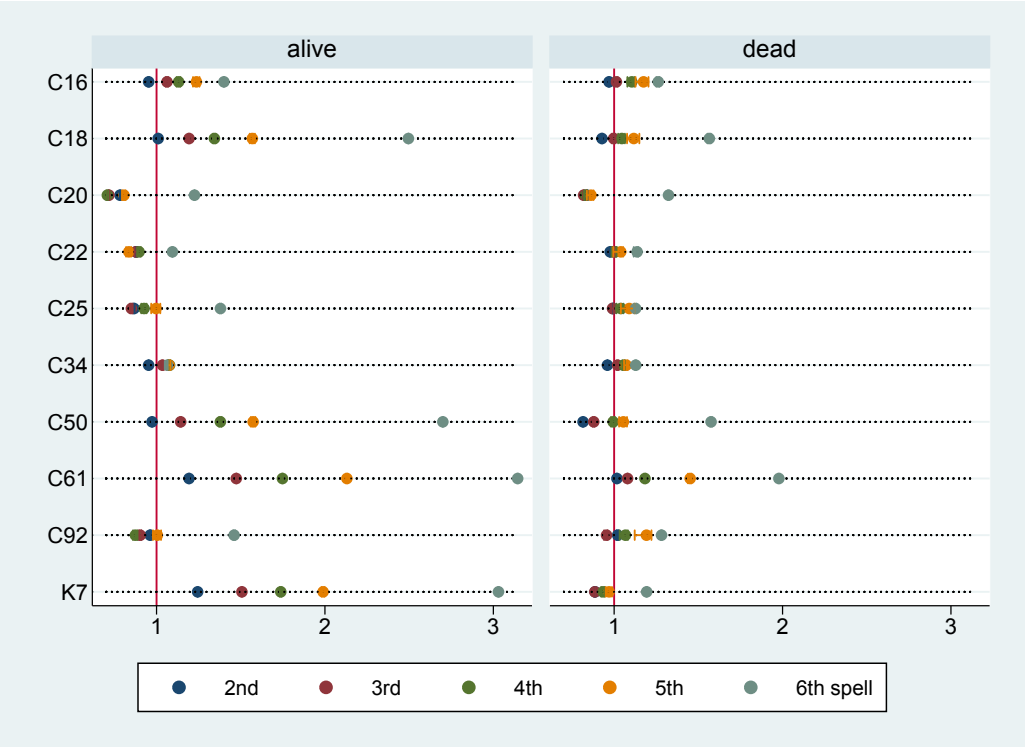


Figure 5: Estimations for the hazard rate shifters of subsequent spells.

Decoding of ICD-10 keys. C16: Malignant neoplasm (MN) of stomach, C18: MN of colon, C20: MN of rectum, C22: MN of liver and intrahepatic bile ducts, C25: MN of pancreas, C34: MN of bronchus and lung, C50: MN of breast, C61: MN of prostate, C92: Myeloid leukaemia, and K70-K77: Diseases of liver.

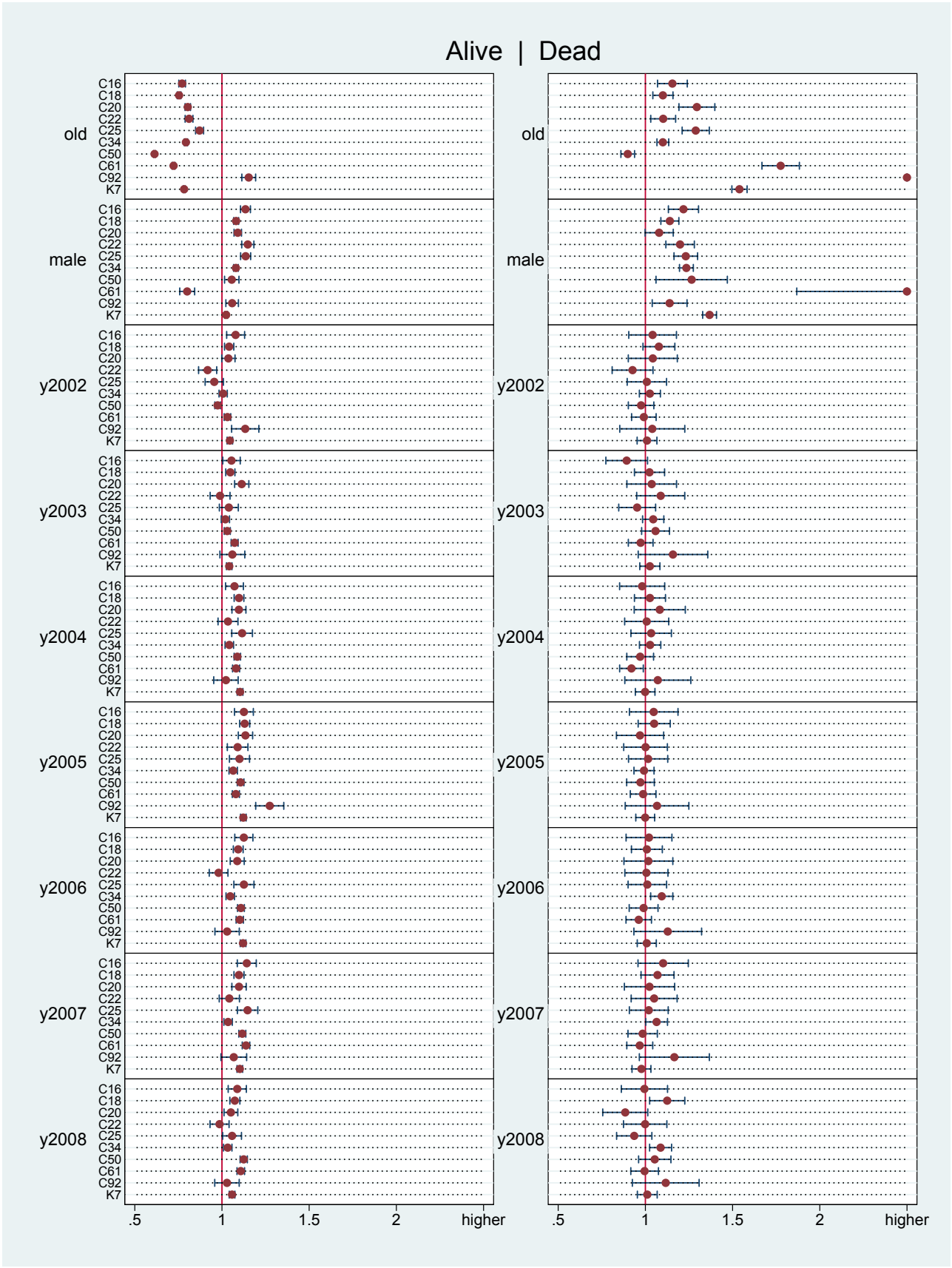
Figure 4*(a)* depicts the hazards that end the first hospital spell during each of the three intervals. The left plot contains the hazard rates for leaving the hospital alive ($\lambda_{i,1}^{IH}$); the plot on the right-hand side shows the baseline hazard rates for dying during the first hospital spell ($\mu_{i,1}^{IH}$). The hazard rate for leaving the hospital alive is clearly decreasing in the spell length for all diseases, whereas the propensity to die is remarkably stable. For all diseases, dying is less likely then being discharged as is indicated by the absolute value of the hazard rates. Note that the hazard rates are estimated quite accurately, as can be verified by the narrow 95% confidence bands for most of the estimates (the bars in the Figure). All hazard rates are (significantly) greater than zero.

The two panels in Figure 4*(b)* show the hazard rates for the competing risks in the first out-of-hospital spell. Again, the left of these panels describe the hazards to leave the out-of-hospital spell alive (i.e. to be rehospitalized, labeled by $\lambda_{i,1}^{OH}$); and the right graph contains the out-of-hospital mortality ($\mu_{i,1}^{OH}$). The absolute value of any hazard is much smaller when compared to the in-hospital hazards illustrated in *(b)*: This mirrors the much longer out-of-hospital spells. The estimates also indicate a decreasing hazard of being rehospitalized over time. The pattern of the hazard rates for dying out-of-hospital, however, is different: hazards start very low in the first interval, then increase substantially in the second interval (relatively speaking), only to decrease to an intermediate level in the third interval.

We now turn to the consequences of stratifying our data for the sequence number of the respective in- or out-of-hospital spell to allow for a sort of history dependence by explicitly allowing the transition patterns to differ across spells. The fact that this assumption is important can be verified when inspecting Figure 5, which contains the estimates of the hazard rate shifters for the subsequent spells. The structure of this figure is the analogous as for Figure 4: Graph *(a)* depicts in-hospital spells with regard to the shifters for ending the second, third, ..., sixth in-hospital spell alive (the left-hand panel) or dead (the right-hand panel); while graph *(b)* depicts the shifters for leaving the subsequent out-of-hospital spells alive or dead. A value of one means that the hazard rates for a spell are the same as for the first spell, and a value below one (above one) implies lower (higher) hazards in the respective spell when compared to the first spell. The fact that stratification is important can be seen from most hazard rate shifters being different from 1.

Consider, for example, breast cancer (C50). For the hazard rates to end one of the subsequent in-hospital spells (i.e., with $l > 1$), the left plot of graph *(a)* reveals that the hazard rates to be discharged alive are relatively constant (the plot indicates shifters between 0.8 and 1.1). However, the hazard rate of breast cancer patients to die during a subsequent spell is increasing and doubles for the sixth spell, as can be verified from the right plot in graph *(a)*. Without stratification we would estimate a mean exit rate for all spells and that would, in the case of breast cancer, bias the estimates of the baseline hazard (i.e., the hazard of the first spell) upward.

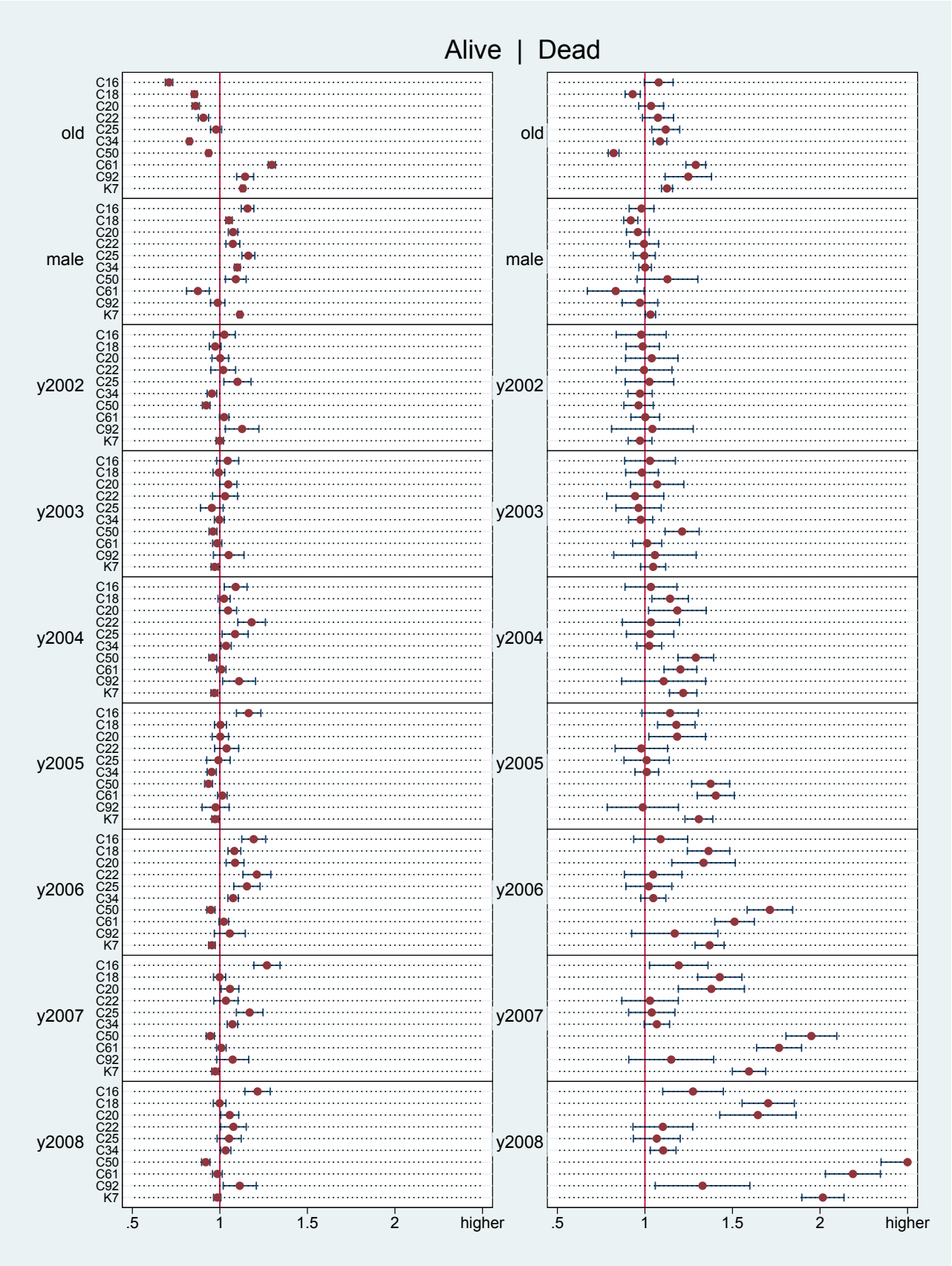
The effect of the covariates is summarized in Figures 6 and 7. Refer to the former: While in-hospital, being old imposes a strong negative effect on the hazard rate of being discharged



(a) discharge: $\boldsymbol{\beta}$ *(b)* death: $\boldsymbol{\gamma}$

Figure 6: Effect of covariates on the hazard rates for in-hospital spells.

Interpretation. Parameter values are multiplied with the corresponding hazard rate. A level of 1 therefore means no change in the hazard rate, a level below (above) 1 implies a lower (higher) hazard rate if the dummy is 1.



(a) rehospitalized: $\boldsymbol{\delta}$ *(b)* death: $\boldsymbol{\eta}$

Figure 7: Effect of covariates on the hazard rates for out-of-hospital spells.

Interpretation. Parameter values are multiplied with the corresponding hazard rate. A level of 1 therefore means no change in the hazard rate, a level below (above) 1 implies a lower (higher) hazard rate if the dummy is 1.

alive whereas it increases the hazard rate of dying. This is true for all diseases (except for C92 and C50). It translates into a longer expected spell-length with a higher probability of dying during hospitalization. The former effect is driven by the relatively high discharge rate compared with the death rate, while the latter effect is simply determined by the sign of the coefficient of the dummy 'old'. Males tend to have shorter stays but exhibit a higher risk of dying. The cohort-year effects reveal (increasingly) shorter stays, but do not identify a significant reduction of in-hospital mortality (beyond what is implied by the higher hazard rates for being discharged alive). Turn now to Figure 7, which summarizes the effect of the covariates on the out-of-hospital spells. Old patients tend to have a lower hazard rate for being rehospitalized and a higher hazard rate for dying out of the hospital, whereas the gender effect reduces to slightly higher hazard rates for rehospitalization (exceptions are C61, C92, and K70-K77). The year of the first diagnosis reveals that no effect on the rehospitalization hazard rates can be identified, but that the later cohorts tend to have a higher risk of dying.

4.3 Multistate model: Survival curves

To obtain disease-specific survival curves, we employ the algorithm of Blaser et al. (2015) as described in Section 2.3, which requires the use of all the estimated hazard rates and coefficients obtained in the previous section. We model a total of 10 in-hospital and 10 out-of-hospital spells. That is, we start with an initial patient cohort and let individuals belonging to the cohort be subject to up to 20 nested competing risks experiments. For the specification of the matrix of the hazard functions, we set transition hazards for spells larger than or equal to 6 to the estimated and fitted hazards of the sixth stratum. Compared to (17), the hatted values now indicate that the hazards are fitted:

$$
\mathrm{H}(t) = \begin{array}{c} \\ IH_1 \\ OH_1 \\ IH_2 \\ \vdots \\ IH_6 \\ OH_6 \\ \vdots \\ IH_{10} \\ OH_{10} \end{array}
\begin{array}{c}
\begin{array}{cccccccccc} OH_1 & IH_2 & \dots & OH_6 & IH_7 & \dots & IH_{10} & OH_{10} & D \end{array} \\
\begin{pmatrix}
\hat{\lambda}_1^{IH}(t) & 0 & \dots & 0 & 0 & \dots & 0 & 0 & \hat{\mu}_1^{IH}(t) \\
0 & \hat{\lambda}_1^{OH}(t) & \dots & 0 & 0 & \dots & 0 & 0 & \hat{\mu}_1^{OH}(t) \\
0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & \hat{\mu}_2^{IH}(t) \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & \dots & \hat{\lambda}_6^{IH} & 0 & \dots & 0 & 0 & \hat{\mu}_6^{IH}(t) \\
0 & 0 & \dots & 0 & \hat{\lambda}_6^{OH} & \dots & 0 & 0 & \hat{\mu}_6^{OH}(t) \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & \dots & 0 & 0 & \dots & 0 & \hat{\lambda}_6^{IH}(t) & \hat{\mu}_6^{IH}(t) \\
0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & \hat{\mu}_6^{OH}(t)
\end{pmatrix}
\end{array}
\tag{24}
$$

The cohort characteristics as described by the covariate vector $\boldsymbol{X}$ are replicated for the simulated cohorts to comply to the cohorts' averages. The size of the artificial cohort is set to equal the size of the observed cohort from the MedStat. Finally, parameter uncertainty is reflected by additionally specifying variance-covariance matrices for each of the non-zero entries of (24) as described in the model section above. In the interest of space, we dispense

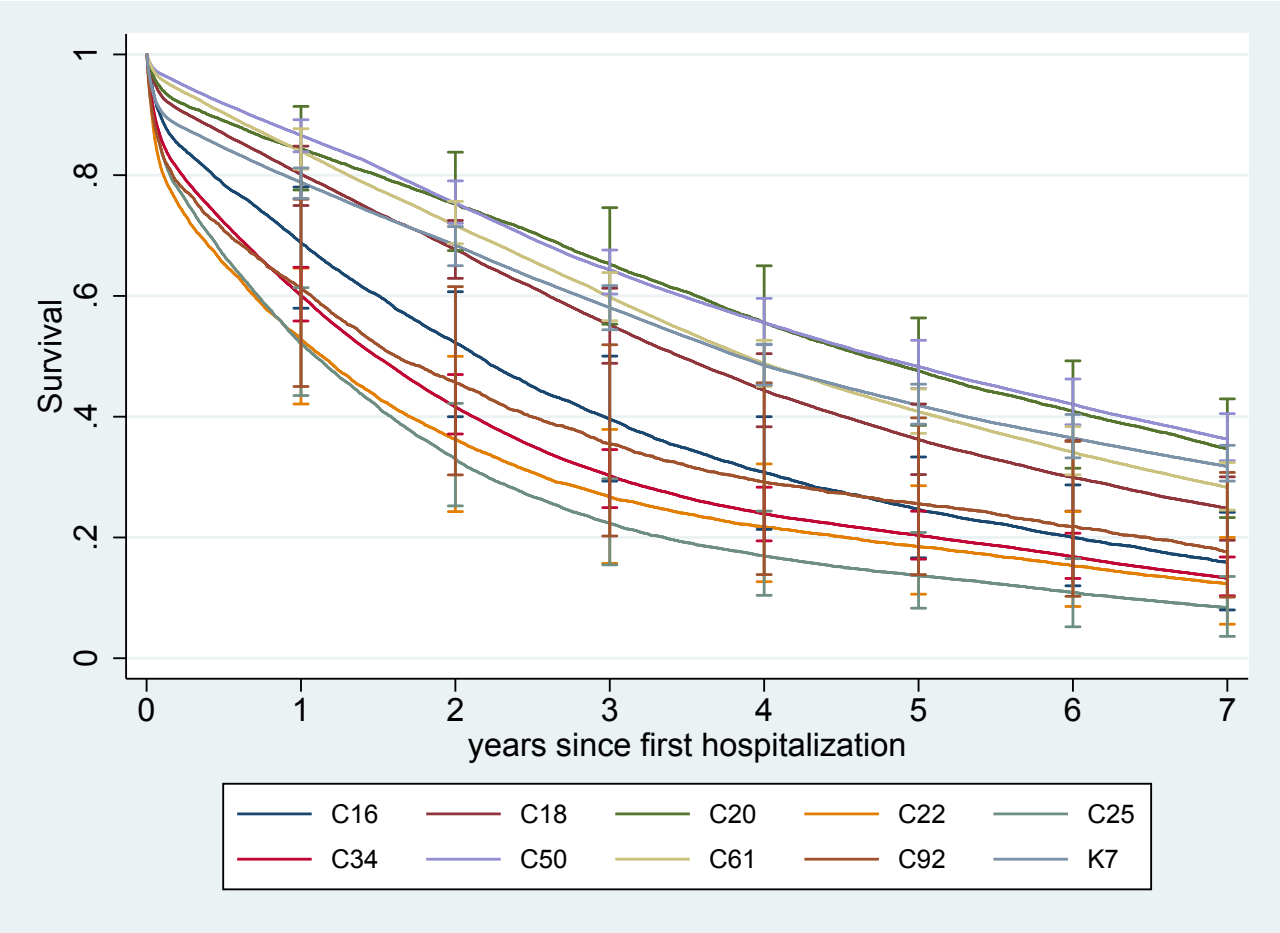


Figure 8: 7-year survival after treatment of chronic diseases in Swiss hospitals (2001-2008 cohorts).

Decoding of ICD-10 keys. C16: Malignant neoplasm of stomach, C18: Malignant neoplasm of colon, C20: Malignant neoplasm of rectum, C22: Malignant neoplasm of liver and intrahepatic bile ducts, C25: Malignant neoplasm of pancreas, C34: Malignant neoplasm of bronchus and lung, C50: Malignant neoplasm of breast, C61: Malignant neoplasm of prostate, C92: Myeloid leukaemia, and K70-K77: Diseases of liver.

with the presentation of the extensive versions of this variance-covariance matrices. Equipped with these estimation results, we simulate the transitions of artificial cohorts in the 21-states multistate model. We do this for every disease and finally compute survival curves, which are depicted in Figure 8.

4.4 Contrasting the multistate and the single-risk model

Figure 9 allows us to compare the predicted survival rates of the multistate model (solid lines) against the prediction of the single-risk model (dashed lines) and the observed survival by NICER (circles). The thin solid and dashed lines depict the 95% confidence bands of the estimated survival rates. We report survival rates for 60 months after the initial hospitalization or diagnosis. Both models reproduce the NICER survival rates relatively well in many cases. When comparing them, however, the multistate model appears to perform worse on average than the single-risk model, although it delivers a superior or at least equally good prediction

for C18, C20, and C92. Especially for the smaller samples, the multistate model has broader confidence bounds. Possible causes and implications of our results are discussed in the next section.

5 Discussion

Looking at the survival times predicted by the multistate model, one cannot conclude that it performs better on average than the prediction of the single-risk approach; in fact, for some diseases it performs worse. That is, taking the hospitalization history of a patient into account does not pay off in terms of an increased accuracy of the prediction. If, however, one is interested in the intermediate transitions and how these are affected differently by the set of covariates, the multistate model might still be the model of choice. The question occurs why the multistate model does not increase the precision compared to the single-risk model, despite the use of more information.

As a general rule for the proposed multistate model, the selected diseases require (1) a sufficiently high mortality and (2) a sufficiently high rehospitalization rate that (3) occur in a sufficiently short time frame. All three criteria were identified heuristically and were used to guide the selection of diseases for which the estimated multistate survival rates are sufficiently close to the NICER observation. We consider nine types of cancer (ICD-10 category C) as well as chronic liver diseases (coded as K70-K77 in ICD-10). This selection covers predominant chronic diseases in terms of mortality (for this and other descriptives, see Table 2). The single-risk approach is less dependent on the aforementioned criteria, which can be a strong argument in favor of it. Overall, Figure 9 reveals a surprisingly good fit of the single-risk model, if we consider the initial doubts on the suitability of this approach for hospital discharge data. A direct implication is that the in-hospital mortality seems to be quite a well suited predictor for overall mortality – at least for our selection of chronic (and quite severe) diseases.

An identifiable disadvantage of our multistate model which we noticed is its relative inflexibility concerning time intervals to divide the out-of-hospital spells. Time intervals are crucial here for theoretical considerations. To illustrate, suppose there was only one time interval over which individuals face two competing risks. The probabilities of an exit into the two states, given a transition occurs, are $\frac{\lambda}{\lambda+\mu}$ and $\frac{\mu}{\lambda+\mu}$, respectively. A direct implication is that the expected time that an individual remains in any given state is equal for both competing risks, and only the share of the population that ends up in either state can be different (and is given by the respective fraction). For our case, this implies that the expected length of an, say, out-of-hospital spell is the same for readmissions or deaths: If (observed) readmissions occur shortly after discharge, so does (and must) death. A way to attenuate this problem is to allow for more than one interval, so that the strict relationship between the expected duration in the two possible states is relaxed. Increasing the flexibility of transition patterns by increasing the number of intervals would improve the estimation, but complicates the convergence of the (out-of-hospital) log-likelihood maximization.

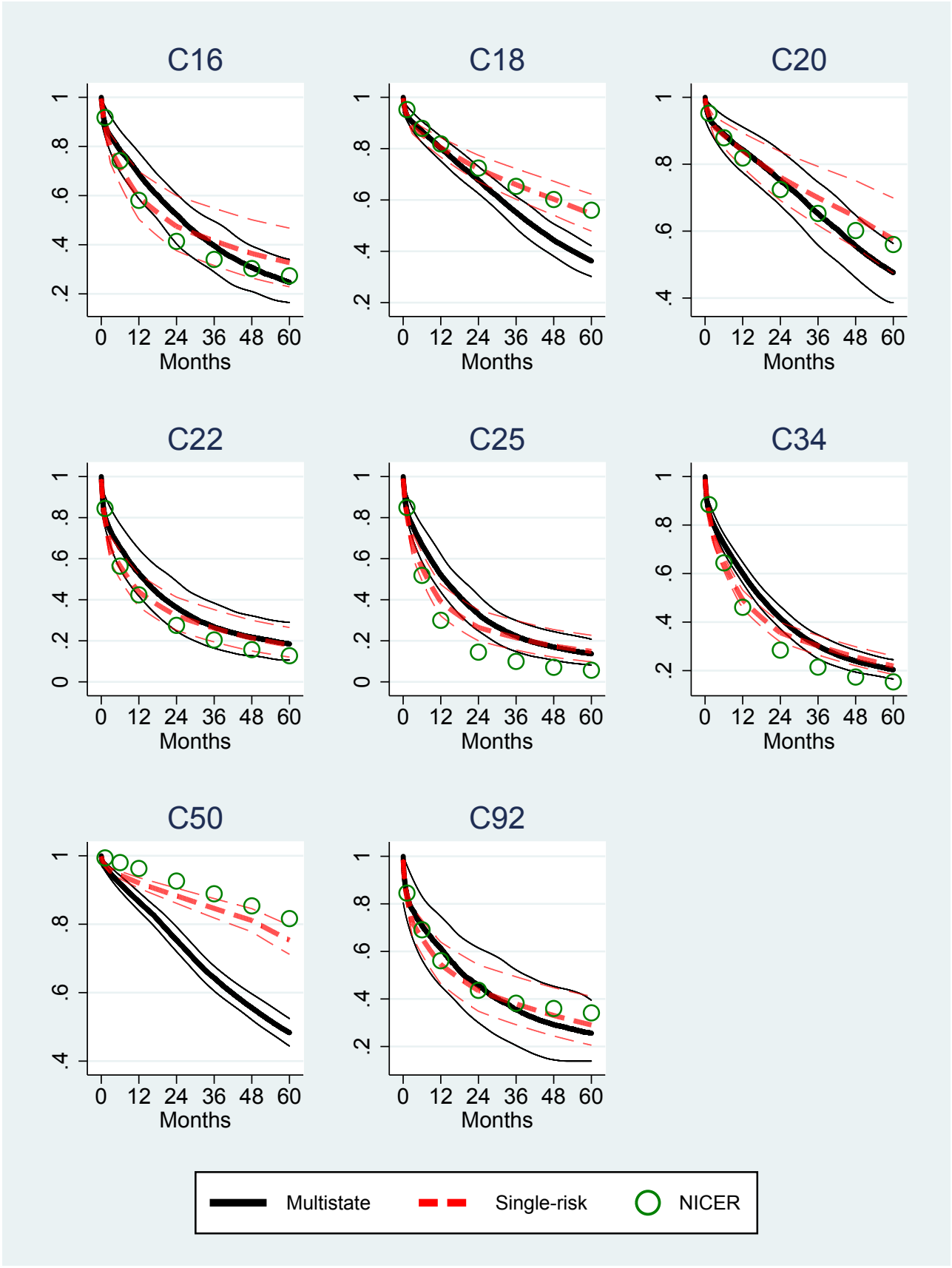


Figure 9: Evaluation of the multistate framework for survival up to 60 months after the initial admission.

y-axis: Survival as a percentage of the initial patient population. *x-axis*: Months after the initial hospitalization for the multistate and naive estimation; months after the initial diagnosis for NICER data.

Indeed, the choice of intervals for the multistate estimation is mainly driven by convergence considerations. These restrictions to the choice of intervals affected both the number of cut-offs and its structure. We found a persistent pattern for all diseases, that the estimation of the competing risks process for out-of-hospital spells only converged for a low number of cut-offs which span over relatively long periods of time. Namely, convergence was attained for three intervals per spell and cut-offs after two and four years (expect C50 where we used 500 and 1,000 days). This implies that the variation in the hazard rates before a total elapsed time of 2 years is mainly captured in the estimates of hazards for in-hospital spells. For in-hospital spells, we generally experienced fewer convergence issues and used the 25% quartile and the median of durations as cut-offs to define intervals. We therefore may not be able to accurately estimate survival rates outside of hospital shortly after discharge. In general, there is a trade-off between the number of intervals (determining the accuracy) and the efficiency of the estimation: more intervals do allow for a more precise account of the time structure underlying the competing risk processes, but come at the cost of estimating a larger number of parameters. This is in particular true for our complex multistate frame, as remember: the number of baseline hazard rates equals $I \times L$, hence an additional time interval leads to L additional parameters being estimated. With our assumption of a proportional shift of hazards for the subsequent spells, an additional interval requires estimation of only one additional hazard; still, we did not achieve convergence for more than three intervals.

To increase our understanding of the source of the inaccuracy, we finally run a substantially simplified version of the multistate model. For this, to estimate the out-of-hospital rates, we only considered at most two admissions of a patient. Thus, we define the first observed admission as onset of risk which runs until the latest point in time that we observe the patient, as in the single-risk case. This point in time is then either the time of discharge or death during the initial admission, if no readmission occurs; or it is the time of death or discharge after the last observed readmission, if a readmission occurred. The process is similar to the single-risk process except that we make use of the information that at least one intermediate hospitalization occurred. The transitions therefore still form a competing risks process, and the out-of-hospital mortality rates were estimated explicitly. The log-likelihood maximization of this model was successful for C16 and C34 with cut-offs at 4 and 80 days for the out-of-hospital process, and the intervals are therefore much shorter than in the previous analysis with multiple out-of-hospital spells. The outcome of this model is depicted with a blue solid line in Figure 10. We observe a somewhat improved fit for early time periods (taking NICER data as the benchmark). One interpretation is that allowing for potential breaks in the hazard function soon after discharge account for the true process more accurately.[19] On the other hand, it appears that this improvement in the fit for small t entails less precision in the fit for larger t, as one can see that the blue solid line diverges from the NICER circles for later months. One must still be cautious in accepting this interpretation, as we were not able to verify the pattern for different or more time intervals. It could therefore also be the case that what we

[19]It is not possible to verify whether there are any structural breaks in the out-of-hospital mortality rates, as the underlying event is not observable.

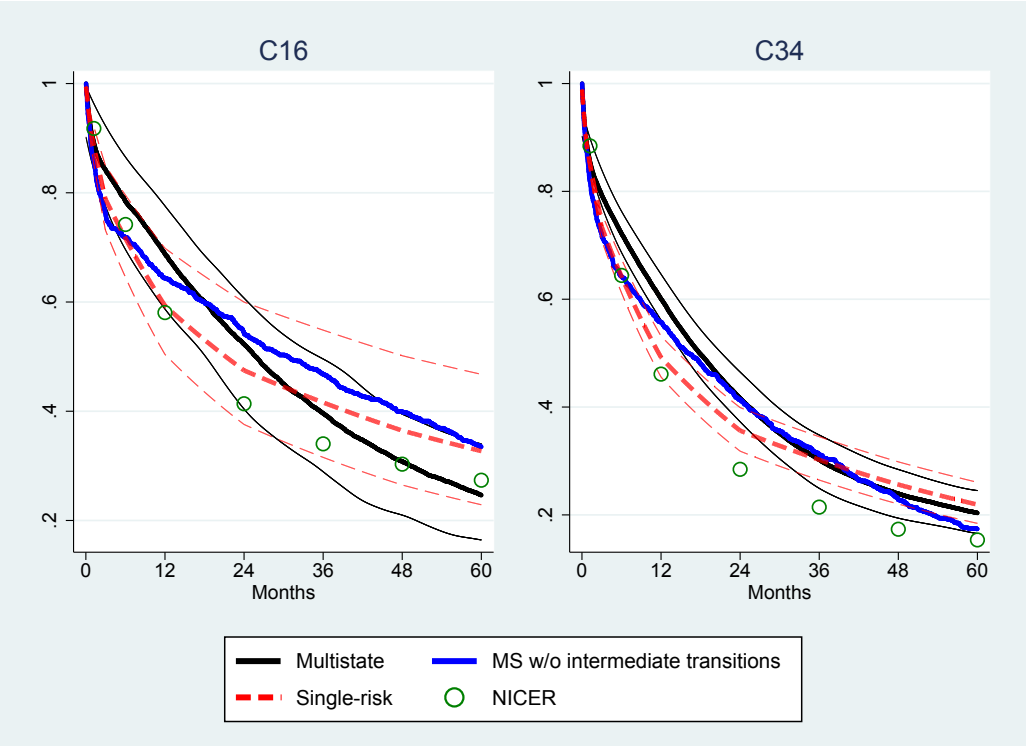


Figure 10: Sensitivity analysis for survival up to 60 months after initial admission.

y-axis: Survival as a percentage of the initial patient population. *x-axis*: Months after the initial hospitalization for the multistate and naive estimation; months after the initial diagnosis for NICER data.

see here is due to a change in the data structure attributable to the disregarded intermediate transitions.

In our set of diseases, breast cancer (C50) sticks out: Mortality is estimated much higher than reported by NICER. In the single-risk case, one could explain such failure with either a violation of the assumption of random censoring or with significant survival improvements over the last decade, which are reflected in the NICER data but not in our cohort data (see below). In the multistate case, the bad performance might come from the combination of an in-hospital mortality rate that is low and a (for the algorithm) misleading pattern of rehospitalizations. Concerning the multistate model, recall that, while out-of-hospital, a patient may be rehospitalized or die, which are the only two transitions possible. For C50, 76% of the survivors are rehospitalized after the first admission, and 50% of these rehospitalizations occur within a period of 6 months after the initial hospitalization (as opposed to 82 % and 1 month for C34). The estimated hazard rates for C50 correctly reflect the share of rehospitalizations, but they fail when it comes to the timing of events. This might be considered as an example of the insufficient flexibility of the intervalization in the multistate model, as discussed above. Alternatively, we might conclude that the typical history of a breast cancer patient no longer fits our implicit definition of a severe and chronic disease, but might more accurately be described as an acute disease.

The assessment of how well both models perform also relies crucially on the assumption that NICER reports the true survival rates. This assumption might be challenged for two reasons

in particular. First, the NICER data are incomplete as they currently only comprise data from 21 of the 26 Swiss cantons. This implies that survival figures for the excluded quarter of the Swiss population are extrapolated by NICER. Though regional variation in the incidence and the course of diseases might exist, we believe that the relative homogeneous populations and health systems across cantons limit this potential source of bias.

Second, the reported numbers from NICER have not been calculated using the cohort approach but using the period approach. In our simulation, however, we used the cohort approach. That means, we follow a cohort of patients after their diagnosis and simply observe when they die, where a cohort of patients consists of all the patients first diagnosed in a specific year. The period approach differs from the cohort approach in the following way: Instead of following a cohort of patients, it considers deaths that occur in a specific time period. To deduce survival times, it looks back in history only as far as necessary, which implies that it is tailored to deliver survival time estimates for the latest respective group of patients.[20] We used the cohort approach for two reasons: The cohort approach allows to identify survival improvements over time; and our estimation strategy to identify the rehospitalizations required to have an equal time frame for every patient during which a rehospitalization might happen (see the discussion on page 17). A problem occurs if mortality changes significantly over time, since in that case we would be comparing apples and oranges. For our set of diseases, Table 4 shows that for some diseases significant mortality reductions are indeed observed. This appears to be particularly relevant for the case of breast and colon cancer. For both we find substantial treatment improvements after 2004 which implies that the reported survival curves from NICER are upward biased in these cases as they do not consider the worse survival rates during the first of our cohorts.

Another reason why the survival rates deduced from our data may differ from NICER regards the definitions of the onset of risk. The cancer registries observe patients since their first diagnosis with the respective cancer. This diagnosis might origin from either an ambulatory or an inpatient examination or treatment and determines the onset of the risk. MedStat, on the other hand, naturally only considers diagnoses made during a hospital spell. Survival rates from both data sources, therefore, are only comparable if not too much time elapses between the patient's first diagnosis and his or her first hospitalization that also reports that diagnosis. Whether or not this is the case depends on the disease and how it is usually treated.[21]

[20] For the present paper, NICER delivered data for the period 2008–2013. To determine survival after, say, 4 years, patients diagnosed between 2004 and 2009 were observed within the 2008–2013 time frame. To determine 2-year survival, no patient diagnosed before 2006 was considered, for 1-year survival no patient diagnosed before 2007 was considered, and so on. Our cohort approach in fact uses 'older' survival times: one-year survival implies that we observed patients diagnosed between 2001 and 2008 who died in the consecutive year; i.e., we did not consider the mortality of patients who died in 2010 or later after being diagnosed for 1 year.

[21] In the case of breast cancer, for example, the clinical guidelines stipulate a surgery (implying a hospital admission) after the diagnosis is confirmed (cf. `http://www.senologie.ch/images/pdf/2003-38-368.pdf`, last accessed June 2018).

6 Conclusion

We compare two possible methods for estimating survival times from hospital discharge data: A single-risk model and a multistate model that takes intermediate transitions into account. For the latter, we extend the approach by Farsi & Ridder (2006) for a multiple of hospitalization sequences. We employ the multistate model to estimate spell-specific hazard rates which we use as input to simulate predictions of survival times based on an algorithm from Blaser et al. (2015). Finally, we contrast the predicted survival times of both the single-risk and the multistate model with observational data from the Swiss cancer registry.

We find that the survival time estimates based on the multistate model are not superior in accuracy to the estimates of the simpler single-risk model. The overall performance of the multistate model depends on the structure of the hospital discharge data. Our estimates approximate the NICER data reasonably well, provided that the diseases at hand exhibit (1) a sufficiently high mortality rate, (2) a sufficiently high rehospitalization rate, and (3) the transitions occur in a short time frame. This precludes diseases which have a mild course.

If, however, one is interested in the intermediate transitions and how these are affected differently by the set of covariates, the multistate model can still be useful. For example, the model could be applied to study the effect of interventions at the patient level, not only on mortality, but also on readmissions after a specific hospitalization spell. The multistate model's application to such inquiries may therefore provide guidance regarding the hospital resource evaluation, the efficacy of certain health policies, and offer policy makers help when planning hospital capacities.

References

Ayanian, J. Z., Kohler, B. A., Abe, T. & Epstein, A. M. (1993), 'The relation between health insurance coverage and clinical outcomes among women with breast cancer', *The New England Journal of Medicine* **329**(5), 326–331.

Beyersmann, J., Allignol, A. & Schumacher, M. (2012), *Competing Risks and Multistate Models with R*, Use R!, Springer Science+Business Media, LLC, New York, NY.

Blaser, N., Salazar Vizcaya, L., Estill, J., Zahnd, C., Kalesan, B., Egger, M., Gsporner, T. & Keiser, O. (2015), 'gems: An R package for simulating from disease progression models', *Journal of Statistical Software* **64**(10), 1–22.

Bucholz, E. M., Butala, N. M., Ma, S., Normand, S.-L. T. & Krumholz, H. M. (2016), 'Life expectancy after myocardial infarction, according to hospital performance', *The New England Journal of Medicine* **375**(14), 1332–1342.

Danieli, C., Remontet, L., Bossard, N., Roche, L. & Belot, A. (2012), 'Estimating net survival: The importance of allowing for informative censoring', *Statistics in medicine* **31**(8), 775–786.

Daysal, N. M. (2012), 'Does uninsurance affect the health outcomes of the insured? Evidence from heart attack patients in california', *Journal of Health Economics* **31**(4), 545–563.

Evans, W. N. & Kim, B. (2006), 'Patient outcomes when hospitals experience a surge in admissions', *Journal of Health Economics* **25**(2), 365–388.

Farsi, M. & Ridder, G. (2006), 'Estimating the out-of-hospital mortality rate using patient discharge data', *Health Economics* **15**(9), 983–995.

Laudicella, M., Li Donni, P. & Smith, P. C. (2013), 'Hospital readmission rates: Signal of failure or success?', *Journal of Health Economics* **32**(5), 909–921.

Martini, G., Berta, P., Mullahy, J. & Vittadini, G. (2014), 'The effectiveness–efficiency trade-off in health care: The case of hospitals in Lombardy, Italy', *Regional Science and Urban Economics* **49**, 217–231.

Meira-Machado, L., de Una-Alvarez, J. & Cadarso-Suarez, C. (2006), 'Nonparametric estimation of transition probabilities in a non-Markov illness-death model', *Lifetime Data Analysis* **12**(3), 325–344.

Rahman, M., Norton, E. C. & Grabowski, D. C. (2016), 'Do hospital-owned skilled nursing facilities provide better post-acute care quality?', *Journal of Health Economics* **50**, 36–46.

Scharfstein, D. O. & Robins, J. M. (2002), 'Estimation of the failure time distribution in the presence of informative censoring', *Biometrika* **89**(3), 617–634.

Wang, M.-C., Qin, J. & Chiang, C.-T. (2001), 'Analyzing recurrent event data with informative censoring', *Journal of the American Statistical Association* **96**(455).

Willekens, F. J. & Putter, H. (2014), 'Software for multistate analysis', *Demographic Research* **31**, 381–420.